An inducible and reversible genetics platform for novel drug target discovery

Leong, Mei Ling

2019

Leong, M. L. (2019). An inducible and reversible genetics platform for novel drug target discovery. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/87335>

<https://doi.org/10.32657/10220/48048>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**An Inducible and Reversible Genetics Platform for
Novel Drug Target Discovery**

Leong Mei Ling

SCHOOL OF BIOLOGICAL SCIENCES

2019

An Inducible and Reversible Genetics Platform for Novel Drug Target Discovery

Leong Mei Ling

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for
the degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

28th February 2019

.....

Date

.....

Ms Leong Mei Ling

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

28th February 2019

.....

Date

Prof Klaus Erik Karjalainen

Authorship Attribution Statement

Please select one of the following; *delete as appropriate:

*(A) This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

~~*(B) This thesis contains material from [x number] paper(s) published in the following peer reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.~~

28th February 2019
.....

Date

.....
Ms Leong Mei Ling

Acknowledgements

Undertaking my PhD at Nanyang Technological University has been both stimulating and enriching. I would like to thank the following people who have made this experience such an enjoyable one:

Prof Klaus Karjalainen for his mentorship and invaluable expert advice. I am ever so grateful that he has given me the freedom to cultivate my ideas.

Prof Christiane Ruedl for her kind concern and suggestions on how I can improve my work.

Dr Piotr Tetlak who showed me the ropes around the lab and was actively involved in teaching and guiding me through the mouse transgenesis work.

Dr Nick Barker for allowing to learn how to culture organoids from his lab at The Institute to Medical Biology (IMB, A*STAR) and Dr Aliya Fatehullah who showed me the technique herself.

Assoc Prof Su I-Hsin and Asst Prof Francesc Xavier Roca Castella, who gave me constructive criticism during the Thesis Advisory Committee meetings.

In my daily work, I have also been blessed with pleasant and helpful colleagues. I would like to thank them for their easy companionship through the years.

Last but not least, I am grateful to my family, especially my parents for their unwavering moral support and encouragement despite having to put up with my eccentricities.

Abstract

Understanding gene function provides insights into biological processes and disease pathologies. Such knowledge can be applied to benefit human health and medicine. With this in mind, I have established a genetic platform in mouse to study gene function. The platform consists of two transgenic mouse strains that were generated in-house, through the use of CRISPR/Cas9 gene editing technology: (1) a strain that ubiquitously expresses the reverse tetracycline trans-silencer (rTetR-KRAB) in all tissues and (2) a strain where the tetracycline response element (TRE) has been inserted in close proximity to the target gene's promoter. With appropriate double transgenic mouse strains, rTetR-KRAB binds to TRE in the presence of doxycycline and represses the target gene. Repression is dose-dependent and reversible. To demonstrate the feasibility of the platform, I have chosen to target *Tcf7l2* which is a transcription factor that acts downstream of Wnt signalling. The Wnt signalling pathway has been implicated in colorectal cancer, but there are contradictory views on whether *Tcf7l2* acts as an oncogene or tumour suppressor gene in this context. Preliminary data obtained using small intestine organoids derived from double transgenic mouse, show the importance of *Tcf7l2* for intestinal homeostasis and ApcMin-associated adenoma maintenance. To expound this, I will utilise the *Tcf7l2* transgenic strain to investigate if physiological rebalance of this gene has therapeutic effects in colon cancer. This platform can also be generally useful to study the physiological or pathological role of other target genes.

Keywords: *tetracycline-inducible gene regulation, CRISPR/Cas9 gene editing, Tcf7l2, Wnt signalling pathway*

Table of Contents

Statement of Originality	3
Supervisor Declaration Statement.....	5
Authorship Attribution Statement.....	7
Acknowledgements.....	9
Abstract.....	11
Table of Contents	13
List of Figures.....	17
List of Tables	18
List of Appendices.....	19
List of Abbreviations	20
Introduction	21
1.1. Studying Gene Function.....	21
1.1.1. Motivations for studying gene function	21
1.1.2. Forward Genetics.....	22
1.1.3. Reverse Genetics	25
1.1.4. Studying gene function <i>in vitro</i> , <i>in vivo</i> or <i>ex vivo</i> ?	26
1.1.5. Manipulating gene expression <i>in vivo</i>	27
1.2. Transgenic Mouse for Animal Studies	28
1.2.1. The mouse as a model organism	28
1.2.2. Achieving spatial and temporal control over gene expression.....	30
1.2.3. Tet-expression system for inducible and reversible gene regulation.....	35
1.2.4. Mouse transgenesis techniques.....	41
1.2.5. CRISPR/Cas9 technology for efficient genetic engineering.....	48

1.3. Tcf7l2 as a candidate gene for study.....	52
1.3.1. Tcf/Lef proteins and Wnt signalling.....	52
1.3.2. Wnt signalling and Tcf7l2 in intestinal homeostasis.....	55
1.3.3. Is Tcf7l2 an oncogene or tumour suppressor in colorectal cancer?	58
1.3.4. Available transgenic mouse strains for studying Tcf7l2 gene function	59
1.4. Establishing a Platform for Novel Drug Target Discovery	62
1.4.1. The proposed platform.....	62
1.4.2. Aims and objectives	63
Materials and Methods.....	67
2.1. Mouse Transgenesis.....	67
2.1.1. sgRNA design	67
2.1.2. Assembly of components for CRISPR/Cas9 gene targeting.....	67
2.1.3. Embryonic stem cell culture.....	69
2.1.4. Transient ES cell transfection.....	71
2.1.5. Fluorescence activated cell sorting.....	71
2.1.6. sgRNA validation.....	72
2.1.7. ES cell clone screening.....	73
2.1.8. Blastocyst microinjection	73
2.1.9. Chimera breeding and colony establishment	74
2.1.10. Mouse genotyping.....	74
2.1.11. mRNA expression analysis	75
2.2. Development of A Doxycycline Treatment Regime	77
2.2.1. Immortalisation of bone marrow cells	77
2.2.2. Production of GFP reporter bone marrow cell line.....	79
2.2.3. <i>In vitro</i> doxycycline kinetics assay	80
2.2.4. Generation of bone marrow chimera mice	81
2.2.5. Doxycycline administration <i>in vivo</i>	81
2.2.6. Determination of GFP repression in bone marrow chimeras.....	81

2.2.7. Tcf7l2 expression analysis in doxycycline-treated mice	83
2.3. Investigation of <i>Tcf7l2</i> function in small intestine organoids	83
2.3.1. R-spondin1 production and purification	83
2.3.2. Organoid derivation	84
2.3.3. Organoid maintenance	85
2.3.4. Assaying function of Tcf7l2 in organoids	86
Results	87
3.1. Creating Transgenic Mouse Strains	87
3.1.1. General plan for creating transgenic mouse strains	87
3.1.2. Identifying sgRNAs that are associated with specific nuclease activity	89
3.1.3. CRISPR/CAS9-mediated gene targeting in embryonic stem cells	93
3.1.4. Generating germline-transmitting chimeras from targeted embryonic stem cells	95
3.1.5. Producing the ROSA26-rTetR-KRAB(Kid1) mouse strain	98
3.1.6. Producing the Tcf7l2-TRE mouse strain	102
3.2. Developing a Doxycycline Treatment Regimen	107
3.2.1. Induction kinetics of rTetR-KRAB(Kid1) <i>in vitro</i>	107
3.2.2. Developing a doxycycline treatment regimen for mice	109
3.2.3. Repression of <i>Tcf7l2</i> in double transgenic mouse	113
3.3. Role of Tcf7l2 in the small intestine	116
3.3.1. Tcf7l2 is essential for the maintenance of small intestine organoids	116
3.3.2. Repression of Tcf7l2 rescues cystic APC ^{min/-} small intestine organoids	117
Discussion	123
4.1. A Platform for Inducible Gene Repression in Mice	123
4.1.1. A choice between Tet-inducible activation or repression	123
4.1.2. Optimising the Tet-inducible trans-silencer for <i>in vivo</i> gene repression	125
4.1.3. Potential Applications of this Platform	127

4.2. Considerations for Creating Transgenic Mouse	128
4.2.1. A pipeline for creating transgenic mouse strains	128
4.2.2. Selecting an assay for screening potential sgRNAs	128
4.2.3. Quantifying sgRNA efficiency	129
4.2.4. Minimising off-target effects	130
4.2.5. Skewing cellular repair pathway for preferential DNA insertion.....	131
4.3. Considerations for Developing a Doxycycline Regimen	132
4.3.1. Pros and cons of the reporter bone marrow chimera system.....	132
4.3.2. Balancing multiple factors to find a suitable doxycycline regimen.....	134
4.3.3. Doxycycline sensitivity and dosage.....	135
4.4. Understanding the role of Tcf7l2 in the intestine.....	136
4.4.1. Relevance of the inducible and reversible Tcf7l2 knockdown mouse.....	136
4.4.2. The role of Tcf7l2 in intestinal homeostasis	138
4.4.3. The role of Tcf7l2 in intestinal cancer	141
4.4.4. Targeting the Wnt signalling pathway through Tcf7l2 as a possible cancer therapy.....	142
4.4.5. Transcription factors as druggable targets	145
Conclusions and Future Perspectives	147
References	149
Appendix A	177
Appendix B	185
Appendix C.....	191
Appendix D	211

List of Figures

Figure 1.1.2. The forward and reverse genetics approaches for studying gene function.

Figure 1.2.2. Transgene systems for spatial and temporal gene regulation.

Figure 1.2.3. Tetracycline-inducible expression system.

Figure 1.2.4. Mouse transgenesis techniques.

Figure 1.2.5. The CRISPR/Cas9 system for genetic engineering.

Figure 1.3.1. The canonical Wnt signalling pathway.

Figure 1.3.2. The healthy epithelial of the small intestine and colon.

Figure 1.4.1. The genetics platform for inducible and reversible gene regulation.

Figure 3.1.1. Approach for creating transgenic mouse strains.

Figure 3.1.2. Identifying sgRNAs that are associated with specific nuclease activity.

Figure 3.1.3. Targeting in embryonic stem cells using CRISPR/Cas9 technology.

Figure 3.1.4. Blastocyst microinjection, chimeras and germline transmission.

Figure 3.1.5. Generation of ROSA26-rTetR-KRAB(Kid1) transgenic mouse strain.

Figure 3.1.6. Generation of Tcf7l2-TRE transgenic mouse strain.

Figure 3.2.1. Induction kinetics of rTetR-KRAB(Kid1) *in vitro*.

Figure 3.2.2. Developing a doxycycline treatment regimen for mice.

Figure 3.2.3. Repression of *Tcf7l2* in double transgenic mice.

Figure 3.3.1. *Tcf7l2* is important for growth of small intestine organoids.

Figure 3.3.2. *Tcf7l2* is important for adenoma formation in APC^{min/-} organoids.

List of Tables

Table 1.2.4. Comparison between sequence-specific nucleases for targeted gene editing.

Table 1.3.4. Available Tcf7l2 transgenic mouse strains.

Table 3.1.3. Summary of frequencies obtained at each stage of embryonic stem (ES) cell targeting process.

List of Appendices

Appendix A: Supplementary figures

Appendix B: Oligonucleotide sequences and PCR conditions

Appendix C: Vector maps and sequences

Appendix D: Permission agreements

List of Abbreviations

APC	Adenomatous polyposis coli
Cas9	CRISPR-associated protein 9
ChIP-seq	chromatin immunoprecipitation – sequencing
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
ES cells	embryonic stem cells
HDR	homology directed repair
IP injection	intra-peritoneal injection
KRAB	Krüppel associated box
mRNA	micro RNA
NHEJ	non-homologous end joining
NLS	nuclear localisation signal
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
RE	restriction enzyme
sgRNA	single guide RNA
TALEN	transcription activator-like effector nuclease
Tcf/Lef	T-cell factor/lymphoid enhancer factor transcription factors
Tcf7l2	Transcription Factor 7 Like 2
tracrRNA	trans-activating crRNA
ZFN	zinc finger nuclease

CHAPTER 1

Introduction

1.1. Studying Gene Function

1.1.1. Motivations for studying gene function

Studying gene function is an important part of biomedical research. It allows us to gain fundamental insights into biological processes and disease pathologies. Such knowledge can be applied to benefit human health and medicine. Some potential applications are described in this section to emphasise the significance of studying gene function.

Understanding disease pathology can lead to the identification of potential therapeutic targets for drug discovery and strategies for therapeutic intervention. Two examples of excellent therapeutic targets are cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and anti-programmed cell death protein 1 (PD-1). With an understanding that CTLA-4 and PD-1 function as negative regulators of T cell activation¹⁻³, Allison⁴ and Honjo³ independently devised a conceivable cancer immunotherapy. They planned to activate T cells by removing the CTLA-4 or PD-1 induced brake, so that the activated T cells would attack tumour cells^{3,4}. Their remarkable success in treating mouse tumours with antibodies against CTLA-4 or PD-1 has been translated to human patients with different cancer types⁵⁻⁷. This defines a new strategy for cancer treatment, known as immune checkpoint therapy.

Studying the biology of bacteria, virus and fungi is useful for preventing or treating infections. For instance, the toxicity of antiviral drugs to viruses, and not their human hosts, is minimised by exploiting the biological differences between the two. Acyclovir was the first antiviral drug to display selective toxicity and many subsequent antiviral drugs are modelled after its success⁸. The reasons for acyclovir's selective toxicity, especially against herpes viruses, are three-fold. First, acyclovir is only converted to its monophosphate form in virus infected cells due to the presence of viral thymidine kinase^{9, 10}. This intermediate monophosphate is subsequently converted to the final active triphosphate form, by other cellular enzymes¹⁰. Second, active acyclovir binds and inactivates viral DNA polymerase more readily than host DNA polymerase¹⁰. Third, active acyclovir is preferentially incorporated into viral DNA during replication and not host DNA^{10, 11}.

The risk of developing a disease can be mitigated through preventive measures, if the underlying cause of a disease is known. This is especially true for multi-gene diseases, where the outcome is influenced by environmental factors. For example, obesity is a major risk factor of type 2 diabetes^{12, 13}. Clinical trials have shown that diet and lifestyle changes addressing this risk factor can delay the onset and even prevent type 2 diabetes^{14, 15}.

1.1.2. Forward Genetics

Two commonly used approaches to investigate gene function are forward genetics and reverse genetics. Both approaches have evolved since their conception. This is especially true for forward genetics because the process has been greatly streamlined by

new tools and resources. The forward genetics approach is described in this section while details of the reverse genetics approach can be found in Section 1.1.3.

When taking a forward genetics approach, the objective is to determine the genetic basis of an observed phenotype (Figure 1.1.2). The first step of this process is to generate a mutant population by random mutagenesis. The population is then screened by visual observation or biochemical assays. Once an individual with an interesting phenotype is found, the causative mutation is mapped by positional cloning. Finally the gene is validated either by performing a rescue experiment or altering the gene's expression in wild type organisms to recapitulate the phenotype. The limiting step in forward genetics screens is identifying the causative mutation; in order to narrow the interval within which the mutation exists, a great deal of time and effort is invested in fine mapping^{16, 17}. Three possible ways to get around this limitation are described below.

First, generate the mutant population with transposon-mediated mutagenesis^{18, 19} instead of random mutagenesis. Transposons insert themselves randomly in the genome, hence causing a mutation and tagging the site at the same time¹⁸⁻²⁰. By making use of the tag, the flanking DNA sequence can be efficiently cloned and determined¹⁸. The sequence can then be searched against a genome database to identify the gene.

Second, identify the causative mutation by whole exome sequencing or whole genome sequencing²¹⁻²³. This method is relatively quick and avoids the need for positional cloning. It is feasible because next generation sequencing technologies has made genome sequencing affordable²⁴.

Figure 1.1.2: The forward and reverse genetics approaches for studying gene function.

Forward genetics is the classic approach to studying function and it precedes reverse genetics. When taking a forward genetics approach, the investigator starts with an observable phenotype and seeks to determine the genetic premise. Reverse genetics proceeds in the opposite direction – the investigator starts with a gene and seeks to determine the associated phenotype or the function, in other words. The main steps in each approach are depicted.

Third, generate the mutant population with a CRISPR/Cas9 library^{25, 26} instead of random mutagenesis. With CRISPR/Cas9 technology, the sgRNA sequence can be varied to induce a Cas9-mediated double-stranded DNA break at the desired site^{27, 28}. DNA repair by the non-homologous end joining (NHEJ) pathway creates an indel mutation²⁹, which can be harnessed for gene knockout. For details on the use of CRISPR/Cas9 technology to induce double-stranded DNA breaks, the reader may refer to Section 1.2.5. So far, usage of CRISPR libraries in forward genetics screens has been restricted to a pooled format^{26, 30-32}. This means that a collective of sgRNAs are added to a single well of cells for lentiviral infection. Thereafter, the causative mutation is identified by whole genome sequencing. In theory, an array format should be possible, whereby each well of cells is transfected or transduced with a single sgRNA²⁵. Doing so precludes the task of finding the causative mutation, because the exact DNA break and mutation site is known beforehand.

1.1.3. Reverse Genetics

The objective of reverse genetics is to determine the phenotype associated with a gene of interest (Fig 1.1.2). The first step of the process is to choose a gene. Next, the gene is either repressed or overexpressed in the organism or cell line. The gene's function can then be deduced by determining which cellular pathways or biological processes are affected. Finally, gene function is validated by performing a rescue experiment.

The choice of the gene is shaped by the study goals. For instance, the International Mouse Phenotyping Consortium^{33, 34} and Knockout Mouse Project^{35, 36} seek to assign function to all annotated genes in the mouse genome. Therefore, genes are systematically knocked out and their function is determined using standardised phenotyping assays. On a much smaller scale, a research group may wish to study a biological process or disease. So they may select a gene with predicted function in their area of interest. There are numerous ways to predict gene function and a few examples are listed, as follows: (1) RNA sequencing (RNA-Seq) or microarray to identify genes that are differentially expressed in healthy versus diseased samples³⁷; (2) functional genomics and proteomics maps to pinpoint dysregulated genes in biochemical pathways³⁷; (3) population-based genetic association studies to identify genetic polymorphisms that are linked to a specific trait or disease^{38, 39} and; (4) comparing DNA sequence or protein homology within or across genomes to find paralogues or orthologues with already described function⁴⁰. Information of this, or similar, nature are readily available online and can be combined to for greater credibility when predicting a gene's function.

Once a gene of interest is selected, its expression can be altered at the protein, RNA or DNA level through the use of small molecule inhibitors, RNA interference (RNAi) technology or transgene expression systems respectively. The advantages and disadvantages of the three methods are considered in Section 1.1.5. Essentially, altering gene expression at the DNA level is a robust method to study the function. However it may require technical expertise, especially if the alteration is made in transgenic animals.

1.1.4. Studying gene function *in vitro*, *in vivo* or *ex vivo*?

Both forward and reverse genetics studies can be performed in various settings. The *in vitro* setting, using cell lines derived either from diseased or healthy individuals, is inexpensive and straightforward but does not recapitulate the complexities within an organism. Recently, organoid culture has gained traction for modelling more complex processes *in vitro* because they self-organise into complex structures with distinct cell types, closely resembling physiological conditions^{41, 42}. However they are still cultured in an artificial environment that lack signalling gradients and input from the stroma^{41, 42}. Therefore any findings will still have to be confirmed with other more stringent methods. *Ex vivo* studies using donated whole organs are biologically relevant⁴³ and about as close to clinical trials as possible. However it is unsuitable for systemic diseases because it is restricted to a single organ. Furthermore, samples are limited and it is both technically and logistically challenging to maintain the organs in a viable state⁴³. *In vivo* animal studies are a well-balanced approach in terms of stringency and feasibility. Data from such studies are valuable because the analysis is performed in the

natural setting, under controlled genetic variability and takes into account the intricacies within a whole organism. Animal studies may be more difficult and time consuming compared to *in vitro* studies, but the benefits far outweigh the effort and costs. In fact, animal studies are specifically chosen for pre-clinical trials because they are the closest to human trials as possible.

1.1.5. Manipulating gene expression *in vivo*

There are several ways to manipulate gene expression *in vivo*, to investigate the effect on the disease. At the protein level, the animal can be treated with small molecule inhibitors for reversible inhibition. However other non-target proteins may be inhibited because of the promiscuity of small molecule inhibitors. For example, type I kinase inhibitors typically interact with many different kinases because their target region, which is the ATP binding pocket, is highly conserved among kinases^{44, 45}. At the translational level, mRNA translation or stability can be regulated with anti-sense technology. However the delivery route may need to be optimised for *in vivo* use and only partial knock down can be achieved. However if knock down proves to be therapeutic, drug design could be curtailed because anti-sense RNAs are therapeutic agents themselves⁴⁶. At the transcriptional level, expression can be perturbed by knocking out the gene or using more sophisticated transgene systems for spatial and temporal control. Such systems include the Cre-LoxP, FLP-FRT, Tamoxifen-inducible Cre, Tet-inducible system, or CRISPR interference to name a few. More details of the various transgene systems and their uses are provided in section 1.2.2. This approach would require the creation of transgenic animals.

Off-target effects or non-specific inhibition can occur with any of afore-mentioned ways of manipulating gene expression. This may raise doubts on whether the findings are truly due to perturbation of the intended target. However this problem can be mitigated if control of gene expression is exerted at the DNA level. This is because the transgenic animals can be mated to segregate off-target mutations. This is another reason why studies using transgenic animals are a stringent way to study gene function.

1.2. Transgenic Mouse for Animal Studies

1.2.1. The mouse as a model organism

Commonly used animal models span a large phylogenetic range, including worms, zebrafish, rodents and non-human primates. The mouse is a popular model organism for several reasons. First, there are many parallels between human and mouse biology, in spite of their size difference. For example, both undergo metabolic changes during ageing⁴⁷ and display distinct behaviours such as depression, aggression and anxiety⁴⁸. Second, mice are more cost-effective when compared to other animal models that more closely resemble humans, such as bigger mammals and non-human primates. This is because mice have accelerated lifespans and produce large litters. Third, there is a wealth of knowledge and available resources for mouse work. For instance, the fully annotated genome is accessible on public browsers and functional genome maps allow better understanding of the physiological similarities or differences between mouse and human. Moreover transgenesis techniques are well established for mice, but not so much for other animal species. Therefore the mouse genome can easily be manipulated to investigate the significance of a particular gene.

It is important to recognise that the mouse is not a perfect model organism. Due to the large evolutionary distance between mice and humans, the biology of the two organisms differs substantially. Primates like the macaque and marmoset might be better models because they are evolutionary closer to humans^{49, 50}. However primates are more costly to maintain compared to mice. Some of the biological differences between mice and humans can probably be attributed to the artificial environment in which mice are housed – germ-free or specific-pathogen-free and with specified diets. It has been shown that, unlike their counterparts from the wildness or pet stores, laboratory mice have an immune system that is akin to human infants⁵¹. Laboratory mice not only have a low number of antigen-experienced CD8⁺ T cells (effector memory and central memory T cells), but they are also unable to mount an effective immune response against *Listeria monocytogenes* infection⁵¹. However after exposing laboratory mice to common mouse pathogens by co-housing with pet store mice, the CD8⁺ T cell populations of laboratory mice start to resemble that of adult humans⁵¹. By no means is this analysis comprehensive, since it focuses only on the CD8⁺ T cells in the peripheral blood. However it suggests that genetic differences may not be the sole reason why mouse research findings, particularly in the field of immunology, do not translate well to humans.

Despite these limitations, research studies using mice have provided valuable insights into the biological function of genes in human physiology and diseases. For example, studies in mice have led to the elucidation of the growth conditions required to culture human intestinal organoids in vitro. Lineage tracing in mouse confirmed that Lgr5⁺ cells are the stem cells of the intestinal crypts. Subsequently, culture conditions to maintain

these Lgr5⁺ mouse stem cells were defined. Further refinement led to the ability to establish organoid cultures from a diverse range of human organs that contain Lgr5⁺ stem cells. Since organoids can be established from a small starting material, such as biopsies, this technique overcomes the limited availability of human samples for research studies. Organoids have been derived from a wide range of human tissues, including the gastrointestinal tract, lung, brain, liver and many more. Therefore the mouse model remains a good starting point for functional studies. But any interesting findings should be verified in more appropriate animal models and with human samples.

1.2.2. Achieving spatial and temporal control over gene expression

The simplest way to perturb a gene in mice would be to create a germline knockout. However such mice are limited in their uses. For instance, a germline knock out strain will not be suitable for studying gene function in adulthood and diseases if the candidate gene is essential during development, because the mouse would be embryonic lethal. In addition, the loss of a gene at the early embryo stage could lead to compensatory mechanisms⁵². To get around these problems, transgene expression systems can be employed for spatial and temporal control over gene expression. I will provide an overview of the common mouse transgene systems and their applications in this section (also reviewed by Zhang and colleagues⁵³ or Branda and Dymecki⁵⁴).

For spatial control over gene expression, the Cre/LoxP or FLP/FRT site-specific recombinase systems can be utilised (Fig 1.2.2 A–B). These binary systems consist of the recombinase (Cre or FLP) and its recognition target sites (LoxP or FRT), which display directionality⁵⁴. The recombinase mediates recombination between two recognition sites⁵⁴. For gene knock out, two recognition sites are introduced in the cis-orientation to flank an essential exon (Figure 1.2.2A). Recombination deletes the intervening exon⁵⁵, resulting in frameshift mutation and premature stop codon. For gene activation, a stop regulatory fragment flanked by recognition sites can be inserted into an early intron of the gene (Figure 1.2.2B). The stop sequence prevents transcription from the nearby promoter so the gene is silent, unless the stop sequence is deleted by Cre- or FLP-mediated recombination⁵⁶. Spatial restriction can be imposed by using a tissue-specific promoter to drive expression of the recombinase in certain cells only^{55, 56}. It should be noted that the Cre/LoxP system may show cross-reactivity in mammalian cells or transgenic mice since cryptic LoxP sites occur with a frequency of 1.2 times per megabase, in the mouse genome⁵⁷. In fact, Cre expression can cause significant karyotype abnormalities in mouse embryonic fibroblast cells, even in the absence of exogenous LoxP sites⁵⁸. This underscores the importance of appropriate experimental controls in order to draw relevant conclusion from the data. For example when using a Cre/LoxP knock out mouse to study gene function, the Cre transgenic mouse would be a better control than wildtype non-transgenic mouse. This is because it accounts for potential cross-reactivity between Cre and cryptic LoxP sites.

Figure 1.2.2. Transgene systems for spatial and temporal gene regulation.

(A-B) Cre/LoxP or FLP/FRT site specific recombinase systems for tissue-specific gene knock out or activation. The recombinase (Cre or FLP) mediates recombination between two recognition sequences (LoxP or FRT), so an intervening sequence flanked by these recognition sites will be deleted. By driving expression of the recombinase from a tissue-specific promoter, gene knock out or activation can be restricted to certain tissues. (C-D) dCas9 transcriptional regulators for tissue-specific transcriptional silencing or activation. Catalytically inactive dCas9 binds to the target sequence and the KRAB or VP16 domain represses or activates transcription from the nearby promoter. For spatial control, the dCas9 regulator can be expressed from a tissue-specific promoter. (E-F) CreER^T system for inducible gene knock out or activation. The Cre recombinase is fused to estrogen receptor. The fusion recombinase only translocates into the nucleus, where it can recombine LoxP sites, when it is bound by tamoxifen. Please refer to the text (Section 1.2.2) for more details. This figure is adapted from reference 56 Copyright © 2012 National Academy of Sciences, reference 59 Copyright © 2013 Elsevier Inc., reference 69 Copyright © Humana Press, a part of Springer Science+Business Media, LLC 2009 and with permission from reference 55 (see Appendix D for permission agreement).

The CRISPR/Cas9 system can also be adapted for gene regulation. The most popular usage of the CRISPR/Cas9 system is as a tool for gene editing, as described in Section 1.2.5. In this system, the two-component Cas9-sgRNA complex is able to recognise and cleave its target DNA sequence^{27, 28}. The sgRNA sequence can be varied, in order to specify the target site^{27, 28}. For gene regulation, catalytically inactive dCas9 is fused to either KRAB domain for transcriptional repression (Figure 1.2.2C) or VP16 domain for transcriptional activation⁵⁹ (Figure 1.2.2D). Both endonuclease domains in dCas9 have been mutated to abrogate its ability to cleave DNA^{27, 60}. However dCas9 is still capable of binding its target DNA in an sgRNA-dependent manner^{27, 60}. Hence The dCas9-sgRNA portion serves merely to guide the chimeric transcriptional regulator to its target locus, while KRAB or VP16 exerts transcriptional control over the target gene⁵⁹. Either the dCas9 regulator or sgRNA can be expressed from a tissue-specific promoter for spatial control of gene regulation. The system could potentially be used for high throughput screens since designing sgRNAs to target different locus is straightforward. However dCas9 transcriptional regulators should be used with caution because Cas9-sgRNA complexes bind in an extremely promiscuous fashion. Genome-wide ChIP-Seq analysis of four sgRNAs demonstrated that they bind anywhere from 25 to more than 5,000 off-target sites⁶¹. Despite promiscuous binding, efficient cleavage only occurred at one or two of the off-target sites⁶¹. This disparity exists because complementarity in the seed region of the sgRNA is sufficient for binding, but cleavage requires extensive complementarity between the sgRNA and target DNA^{61, 62}. Conformational studies also show that the HNH nuclease domain of Cas9 only shifts into the correct position to nick the DNA when there are less than four mis-matches between the sgRNA and target DNA⁶³. This is reminiscent of miRNA-mediated gene silencing. If there is full complementarity between the target mRNA and miRNA, the mRNA is cleaved by the

RNA-induced silencing complex (RISC)^{64, 65}. However if there is only partial complementarity, the target mRNA is typically destabilised through loss of the poly(A) tail followed by the 5' cap⁶⁵⁻⁶⁷. A better understanding of sgRNA binding properties will be required to increase the precision of dCas9 regulators, before they can be used effectively.

For temporal control over gene regulation, there is an inducible version of the Cre/LoxP system, known as CreER^T/LoxP^{68, 69} (Figure 1.2.2E-F). It exploits the estrogen receptor's ability to translocate into the nucleus when bound by its natural ligand estradiol or the small molecule inhibitor tamoxifen⁷⁰. Cre recombinase is fused to a mutated ligand binding domain of the estrogen receptor (ER^T), but this chimeric recombinase is inactive because it is localised to the cytoplasm⁶⁹. Administering tamoxifen induces nuclear translocation of the recombinase where it can recombine DNA at LoxP sites⁶⁹. When adapting this system for mice, it is necessary to use the mutated ER^T estrogen receptor, because the receptor has lost its ability to be activated by endogenous estradiol but not tamoxifen⁷¹. In addition to temporal control, spatial control can be achieved by expressing CreERT2 from a tissue-specific promoter. However the outcome, be it gene knock out or activation, is irreversible once the recombination event has happened.

The Tet-inducible expression system also allows temporal control over gene expression, but has the added advantage of being reversible. The reason for this is because it regulates gene expression at the epigenetic level. The Tet-inducible system has come a long way since it was discovered and a multitude of versions have been created. In next section (Section 1.2.3), I will outline the uses of the Tet-inducible system and highlight key improvements that have transformed the original system.

1.2.3. Tet-expression system for inducible and reversible gene regulation

The Tet-inducible system was derived from bacteria, in particular the *Escherichia coli* Tn10 transposon⁷². Tn10 encodes the tetracycline anti-porter protein (TetA), which confers resistance against tetracycline antibiotics. TetA actively transports tetracycline out of the bacteria⁷³ so that the antibiotic is unable to interfere with protein synthesis. The same transposon also encodes the tetracycline-responsive repressor (TetR), for regulation of TetA expression⁷³ (Figure 1.2.3A-B).

In the original adaption of this system for gene regulation in mammalian cells, TetR was fused to the herpes simplex virus VP16 transactivator domain⁷² (Figure 1.2.3C). This chimeric trans-activator was responsive to doxycycline, a tetracycline analogue, because of the TetR domain. In the presence of doxycycline, the trans-activator remained free-floating. However, without doxycycline treatment, TetR underwent a conformational change that allowed the trans-activator to bind the tetracycline response element (TRE), a heptad repeat of TetO. The VP16 component of the trans-activator triggered transcription from the nearby minimal promoter. Instead of using a full promoter, the authors opted for a minimal promoter so as to avoid transcription when the trans-activator is not tethered nearby.

The Tet-inducible system has seen several key developments since its original version.

The most prominent development, and also the first, was to reverse the binding property

of the trans-activator. This was done by randomly mutating TetR and screening for a version that could bind TRE in the presence of tetracycline⁷⁴. The reverse TetR (rTetR) that was identified carried four amino acid mutations and responded to doxycycline in a dose-dependent manner, much like the TetR version⁷⁴. rTetR can be fused to VP16 for transcriptional activation in the presence of doxycycline (Figure 1.2.3E). The rTetR-VP16 is advantageous for *in vivo* applications because doxycycline is only administered when gene activation is desired, unlike the continuous administration that is required with TetR-VP16. This means that induction of the system is quick since there is no need to wait for doxycycline clearance. However rTetR-VP16 displayed residual affinity for TRE in the absence of doxycycline, resulting in leaky expression^{75, 76}. The problem of basal leakiness was probably compounded by the use of the VP16, since it is an extremely potent transcriptional activator^{77, 78}. The solution to this was to lower the basal leakiness through the use of a dual TetR switch system which comprises the TetR-KRAB trans-silencer and rTetR-VP16 trans-activator^{79, 80} (Figure 1.2.3G), further truncating the minimal promoter to remove promoter elements within the leader⁸¹ or mutating rTetR to significantly reduce its basal activity⁷⁶. Since all three methods improve different aspects of the system, they can be combined for even better reduction of basal leakiness. Reducing basal leakiness was seen as the way to tackle this problem rather than using a less potent trans-activator. This is because a less potent trans-activator would lower the level of expression that can be achieved, hence reducing the dynamic range.

Figure 1.2.3. Tetracycline-inducible expression system.

(A-B) Tetracycline resistance in bacteria is granted by the Tet operon. The Tet operon encodes for both tetracycline repressor (TetR) and tetracycline anti-porter protein (TetA). When tetracycline is absent, the system is in the “off” state. TetR binds to the tetracycline operator (TetO) and it represses TetA expression. However when tetracycline is present the system is in the “on” state. TetR undergoes a conformational change when tetracycline interacts with it. This causes TetR to disengage from TetO so that TetA is expressed. TetA localises to the plasma membrane and actively transports tetracycline out of the bacteria cell. At the same time, more

TetR is synthesised so that TetA synthesis can be switched off once tetracycline is cleared from the cytoplasm. **(C-D)** Tet-expression system for inducible and reversible gene overexpression or knock down in the absence of doxycycline (dox). TetR is fused to VP16 or KRAB domains to create a chimeric transcriptional activator or repressor respectively. In the absence of doxycycline, the transcriptional regulator binds to the tetracycline response element (TRE), which is a heptad repeat of the TetO sequence, and either activates (TetR-VP16) or represses (TetR-KRAB) transcription. In the presence of doxycycline, the transcriptional regulator detaches from TRE so gene regulation is relaxed. **(E-F)** Tet-expression system for inducible and reversible gene expression or knockdown in the presence of doxycycline. rTetR is fused to VP16 or KRAB domains to create a chimeric transcriptional activator or repressor, respectively. rTetR is a mutated version of TetR that has reverse binding properties. This means that rTetR binds to TRE in the presence of doxycycline, unlike TetR that binds when doxycycline is absent. Therefore doxycycline is only administered when transcriptional activation (rTetR-VP16) or repression (rTetR-KRAB) is desired. When doxycycline is absent, the transcriptional regulator detaches from TRE so gene regulation is relaxed. **(G)** Dual TetR switch system for tight regulation of gene expression. In the absence of doxycycline, there is leaky expression because the rTetR-VP16 trans-activator retains some ability to bind the tetracycline response element (TRE). However the introduction of the TetR-KRAB trans-silencer abolishes leakiness because it preferentially binds TRE and ensures repression. When doxycycline is added, the trans-silencer disengages from TRE and is replaced by the trans-activator. Hence the trans-activator drives expression from the minimal promoter. Abbreviations: GOI gene of interest, mini minimal promoter. This figure is adapted from reference 73, Copyright © 2016 Bentham Science Publishers and with permission from references 180 and 93 (see Appendix D for permission agreements). Figure G is based on information from reference 79.

Another development was to reduce the 130 amino acid VP16 domain⁷² to a 12 amino acid minimal activation sequence⁷⁴. It has been shown that overexpression of the yeast GAL4 or chimeric GAL4-VP16 transcriptional activators non-specifically inhibit transcription of non-GAL4 target genes and causes cell toxicity^{82, 83}. This is known as squelching and it happens because an overwhelming amount of GAL4 soaks up the general transcription factors, leaving an insufficient quantity of transcription factors for overall transcription⁸². By removing the binding sequences of transcription co-factors such as Oct-1, HCF, TAFII40 & ADA2, the minimal VP16 trans-activator was better tolerated at high intracellular concentrations⁷⁴. This has the added benefit of reducing the length of the transgene construct since only three copies of the minimal 12 amino acid sequence was sufficient to achieve similar a activation level as the full length VP16 domain⁷⁴. The minimal VP16 improvement may be more of a pre-emptive strike because to my knowledge there have not been published examples of squelching by rTetR-VP16 or TetR-VP16.

The Tet-inducible system has mostly been used for overexpression. However ectopic expression or overexpression could cause pleiotropic effects, so the data obtained with such models should be interpreted with caution. Take for example mouse models of Alzheimer's disease. In many Alzheimer's models, amyloid precursor protein (APP) is overexpressed in the brain to provoke A β plaques accumulation, which is one of the characteristics of Alzheimer's disease. When A β plaques accumulate due to APP overexpression, calpain is activated and it cleaves p35 to form p25 and p10⁸⁴. Elevated p25 results in constitutive Cdk5 kinase activation, triggering a sequence of events that

start with hyper-phosphorylation of Tau, followed by formation of neurofibrillary tangles and eventually neurodegeneration⁸⁵⁻⁸⁷. In contrast, it is postulated that Cdk5 kinase activity in healthy individuals is short-lived because the kinase is activated by p35, which is more rapidly degraded than p25^{84, 88}. This finding has excited much furore over p25 as a potential therapeutic target for Alzheimer's disease. However the p25 phenotype might be an overexpression artefact, since it is not observed when formation of A β plaques is induced by mutating the wildtype APP gene into the pathological Swedish and Beyreuther/Iberian versions^{89, 90}. Therefore knock out or repression is preferred for studying gene function. The Tet expression system can be used for inducible and reversible transcriptional repression, by fusing TetR or rTetR to the KRAB transcriptional repressor domain⁹¹⁻⁹³ (Figure 1.2.3D & F). In opposite fashion to the trans-activator, the TetR-KRAB or rTetR-KRAB trans-silencers repress transcription when they are tethered close to the target gene's promoter.

Unlike the commonly used mouse transgene systems described in the previous section, the Tet-inducible expression system allows fine control over gene expression; Not only is it inducible and reversible, but spatial control can also be achieved by expressing the trans-activator or trans-silencer from a tissue-specific promoter. However all transgene systems are cumbersome because two transgenic strains are required. Moreover at least one of the strains has to be a targeted knock in, which is more difficult to manage compared to random integration.

1.2.4. Mouse transgenesis techniques

The research community has long recognised the potential of genetic engineering and has strived to advance this field. Out of all the species, mouse transgenesis techniques are the most advanced, allowing for complex gene manipulations. These techniques are crucial for establishing *in vivo* mouse transgene systems, which requires insertion of exogenous DNA into the genome.

Foreign DNA can be introduced into the mouse genome either directly or indirectly. With the direct method, the purified DNA is injected into the pronucleus of zygotes⁹⁴ (Figure 1.2.4A). The DNA will integrate into a random locus in host genome. Typically, the integration event happens in the single cell zygote stage so the resulting transgenic founder mouse carries the transgene in all cells of the body. However if integration happens at a later stage, the event is different in each cell of the embryo so the resulting transgenic founder mouse is mosaic. The founder mouse can be bred to establish a transgenic mouse colony because the transgene can be transmitted through the germline. Alternatively, exogenous DNA can be packaged into retroviruses and introduced by infection of early morulae^{95, 96}. Only one copy of the exogenous DNA is integrated and at a non-targeted locus⁹⁷. However it was noted that expression of the transgene was silenced as the founder mouse aged⁹⁶. The direct method is typically used to create gain-of-function mouse or to insert an anti-sense RNA expression cassette for gene silencing via RNA interference.

The indirect method requires an extra step, where DNA is first introduced into ES cells before engineered ES cells are injected into blastocysts⁹⁸⁻¹⁰¹ (Figure 1.2.4B). The resulting mouse is a chimera with contribution from both the injected ES cells and host blastocysts. If the injected ES cells formed the germ cells, the transgene modification will be transmitted through the germline. With this method, it is possible to attain gene modification at a target locus in spite of the rarity of such events, because a large number of ES cell clones can be processed simultaneously. To aid the process, a positive-negative selection strategy, which comprises of the neomycin resistance and HSV-thymidine kinase genes, is used to enrich for clones that have the targeted gene modification¹⁰⁰. The applications of this indirect method are widespread. It can be used for gene knock out, gene trapping, gain-of-function and much more. However it is trickier than the direct method, because minor deviations in culture conditions may cause ES cell lines to lose their germline competency. Loss of germline competency is normally due to the acquisition of chromosomal aberrations, for example trisomy 8¹⁰². ES cells with this particular chromosomal abbreviation were not germline competent, yet they overtook the culture because of the growth advantage afforded by this aberration¹⁰². Accumulation of epigenetic abnormalities in ES cells during culture can also affect the number of chimeric founders. This is because the deviant epigenetic marks persist during embryo development and lead to phenotypic abnormalities in chimeric fetuses¹⁰³, possibly affecting the viability of fetuses.

Figure 1.2.4. Mouse transgenesis techniques.

(A) Direct method to create transgenic mouse in one step. Exogenous DNA can be introduced into mouse embryos by two methods. The first method is to inject purified DNA into zygotes,

and allow the DNA to incorporate into the genome by homologous recombination. The second method is to package the DNA in retrovirus and infect early stage morulae with said retrovirus. This exploits the retrovirus's mechanism to insert the DNA into the mouse genome. The manipulated embryos are then implanted into pseudopregnant mothers and carried to full term. Pseudopregnant females were receptive to the embryos because they were mated to vasectomised males prior to implantation, thus stimulating the hormonal cues required for pregnancy. The resulting pups are mosaic and can be bred to establish a transgenic line **(B)** Indirect method to create transgenic mouse in two steps. In the first step, embryonic stem (ES) cells are genetically engineered to carry the desired modification. These cells are then introduced into the blastocyst in the second step. The blastocysts are then implanted into pseudopregnant mothers and carried to full term. The resulting pups are chimeras and can be bred to establish the transgenic line. **(C)** Sequence-specific nucleases for targeted gene editing. With the exception of meganucleases, the other three nucleases can be programmed to create a double-stranded break at a specific locus. Repair by non-homologous end joining leads to indel mutations, which can be used to knock out a gene if positioned in an exon. If knock in is desired, a replacement template with the desired sequence can be supplied. Homology-directed repair will insert the sequence into the target locus. For details of each nuclease and their limitations, the reader may refer to Table 1.2.4.

To increase the efficiency of targeted gene modification, both the direct and indirect transgenesis methods can be supplemented with sequence-specific nucleases including meganucleases, zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) or clustered regularly interspaced short palindromic repeats-

associated protein 9 (CRISPR/Cas9). These nucleases are able to recognise specific sequence and cleave both DNA strands at that locus. This triggers DNA repair, which can be harnessed to either knock out a gene or insert a piece of DNA (Figure 1.2.4C). Repair by the non-homologous end joining (NHEJ) pathway leads to indel mutations at the double-stranded break²⁹. By positioning the break in an exon, indel mutation would lead to a premature stop codon and hence gene knockout. For knock in, a replacement template with the desired insert sequence flanked by homology arms has to be supplied. Repair by the homology-directed repair (HDR) pathway inserts the desired sequence into the target locus²⁹. Each nuclease has its own advantages and disadvantages, as summarised in Table 1.2.4. It is difficult to compare the targeting efficiencies across the four nucleases because efficiency can be affected by factors that have nothing to do with the mechanism itself. For example, both TALEN and CRISPR/Cas9 are affected by chromatin accessibility⁹³, so the same locus has to be targeted in both cases for a meaningful comparison. Even when this variable is kept constant, difference in expression could be attributed to expression levels of the nuclease¹⁰⁴ and transfection efficiency. The best way to control for such variables would be to separately assemble all the required components for each of the nucleases and perform the reaction in a tube. However going to such extremes just to identify the most efficient nuclease is superfluous. The consensus in the research community is that CRISPR/Cas9 is the practical choice since it is efficient, and by far the easiest to design and implement out of all the nucleases¹⁰⁴⁻¹⁰⁶.

Table 1.2.4. Comparison between sequence-specific nucleases for targeted gene editing.

	Meganuclease	ZFN	TALEN	CRISPR/Cas9
Origin	Yeast and algae ^a	Engineered chimeric protein	Engineered chimeric protein ^b	Bacteria ^c
DNA binding domain	Meganuclease	Zinc finger proteins	TAL modules	sgRNA
Cleavage domain	Meganuclease	FokI	FokI	Cas9
Efficiency^d	Not determined	Variable ^e	Moderate	High
Target site length	14-40 bp	18-36 bp	30-40 bp	20 bp + 3 bp PAM
Specifying target site	Extremely difficult. Sequentially mutate DNA interface of meganuclease ²⁷³	Difficult. Advanced knowledge of zinc finger protein-DNA binding properties required.	Moderate. Combine various TAL modular domains, each with a known specificity.	Easy. Watson-Crick base pairing rules.
Off-target effects^f	Not determined	Low	Low	Higher (See section 4.2.4 for ways to reduce off-targets)
Assembly of components	Not required, can be used as it is. Meganucleases are available commercially.	Tedious. Molecular cloning to assemble zinc finger protein into sequence array, before synthesising the protein.	Tedious. Molecular cloning to string together TAL modular domains ²⁷⁴ , before synthesising the protein.	Easy. Molecular cloning to insert variable sgRNA sequence into expression vector. Host cell machinery to express sgRNA and Cas9.

(Notes on next page)

^a The most common meganucleases for gene targeting are I-SceI and I-CreI from *Saccharomyces cerevisiae* (yeast) and *Chlamydomonas reinhardtii* (algae) respectively. However meganucleases are also found in bacteria, archaea, and fungi.

^b TAL effectors are from *Xanthomonas* bacteria.

^c The most common CRISPR/Cas system for gene targeting originates from *Streptococcus pyogenes* bacteria. However they also occur naturally in other bacteria and archaea.

^d Targeting efficiencies are from references 104, 105 and 106

^e Targeting efficiencies are highly varied because designing zinc finger nucleases is a hit-or-miss process.

^f The frequencies of off-target effects are from reference 105

1.2.5. CRISPR/Cas9 technology for efficient genetic engineering

The CRISPR system was discovered as a bacteria defence mechanism to protect against phage infection¹⁰⁷. It has been likened to adaptive immunity because it has the ability to “remember” a previous infection by incorporating pieces of phage DNA into the bacteria CRISPR locus, hence gaining resistance against future infections by a similar phage¹⁰⁷. The ability to guard against subsequent phage infections is due to sequence-specific recognition and cleavage of the foreign DNA¹⁰⁸ by a process known as CRISPR interference. Many variations of the CRISPR system exist, each with distinct properties (also reviewed by Mohanraju and colleagues¹⁰⁹). For instance, type VI CRISPR system cleaves RNA instead of DNA¹¹⁰.

The type II CRISPR/Cas system was adopted for genetic engineering because it is the simplest of the CRISPR systems. It only requires a single Cas protein for DNA cleavage²⁷, unlike the other systems that require multiple Cas proteins¹¹¹. In bacteria, the type II CRISPR locus codes for a total of three Cas proteins: Cas1 and Cas2 for CRISPR acquisition, and Cas9 for CRISPR interference¹¹¹. During acquisition, foreign DNA that has been processed by Cas1 and Cas2, are incorporated into the CRISPR array as protospacer DNA interspaced by short palindromic repeat sequences¹⁰⁹ (Figure 1.2.5A). During interference, active transcription from the CRISPR array produces a long RNA, which is processed into shorter CRISPR RNAs (crRNAs)^{112, 113} (Figure 1.2.5A). Cas9, crRNA and tracrRNA come together to form an active complex, which

recognises its cognate sequence by complementary base pairing between crRNA and the target exogenous DNA and cleaves the DNA¹¹¹. There are two ways in which bacteria ensure that they do not harm themselves. First, the target DNA sequence must be followed by a protospacer adjacent motif (PAM) for Cas recognition and cleavage¹¹⁴. This prevents Cas from cleaving itself at the CRISPR locus because the integrated spacer sequence does not have the required PAM site. Second, bacteria preferentially acquire spacers from foreign DNA rather than their own chromosome. In bacteria, RecBCD is an intrinsic machinery used to repair double-stranded DNA breaks at stalled replication forks. It unwinds and degrades DNA until it reaches a *Chi* site. This generates fragmented DNA, which is a source of DNA for spacer acquisition. The bacteria avoids extensive degradation of its chromosome because it is enriched with *Chi* sites, hence there is less material for spacer acquisition¹¹⁵. However linear phage DNA has a much lower frequency of *Chi* sites, so it is easily degraded by RecBCD and acquired as CRISPR spacers¹¹⁵.

In 2012, two independent research groups showed that the type II CRISPR/Cas9 system could be reconstituted *in vitro* for RNA-guided DNA cleavage, using a three-component system^{27,28}. Subsequently, it was demonstrated that the CRISPR/Cas9 system could be applied for targeted gene editing in mammalian cells^{116, 117} and to generate transgenic mouse with targeted gene modification^{118, 119}. The three-component CRISPR/Cas9 genetic engineering system consists of the Cas9 protein, tracrRNA and crRNA.

Figure 1.2.5. The CRISPR/Cas9 system for genetic engineering.

(A) The type II CRISPR/Cas system in its native setting, in bacteria. The CRISPR locus codes for Cas1 and Cas2 proteins, which are involved in CRISPR acquisition. They process foreign DNA and incorporate them, as spacers, into the CRISPR array. The components required for CRISPR interference, namely Cas9 protein, CRISPR RNA (crRNA) and trans-acting crRNA (tracrRNA), are also expressed from the CRISPR locus. These three components come together to form an active complex that recognises its cognate sequence on foreign DNA and cleaves it. Sequence specificity is conferred by the crRNA. (B) Adaptation of the type II CRISPR/Cas system for genetic engineering. The crRNA and tracrRNA components from the original system were combined into a single guide RNA (sgRNA). The sgRNA carries a 20 nt variable region that determines the target locus. Therefore the user can target different loci by specifying the sequence of this variable region. The target DNA must contain a “NGG” protospacer adjacent motif (PAM) for recognition by Cas9. When these two conditions are met, the Cas9-sgRNA complex cleaves the target DNA to create a blunt end break. Part A of this figure is adapted with permission from reference 111 and part B is reproduced with permission from reference 29 (see Appendix D for permission agreements).

However tracrRNA and crRNA can be combined into a single guide (sgRNA), without adversely affecting gene targeting efficiencies²⁷. Two requirements have to be met for Cas9 to cleave the target DNA (Figure 1.2.5B). First, the target DNA must contain a “NGG” PAM. This is necessary because interaction between PAM and the phosphate lock on Cas9 pries apart the two DNA strands so that sgRNA can invade and form a R-loop with the DNA^{114, 120, 121}. Once this happens, the second requirement comes into play: the 20 nt variable region of the sgRNA must match the target DNA sequence. If there is only partial complementarity between the sgRNA and target DNA, cleavage will not occur⁶¹⁻⁶³. If there is high complementarity, Cas9 undergoes a conformational change that puts the two nuclease lobes in position to nick both DNA strands^{63, 121}, creating a blunt end double-stranded break. The CRISPR/Cas9 system is a highly efficient technique for genetic engineering. Targeted knockout mice have been generated via direct zygote microinjection, with efficiencies ranging from 47% to 100%¹²²⁻¹²⁴. Targeted insertion is more challenging, but still feasible. A targeting efficiency of 8 - 10% was obtained when inserting a 700 bp fragment in the Nanog and Oct4 gene loci¹¹⁹.

The PAM requirement may present a challenge when precise gene editing is required because it limits the target sites. For example, trying to change a single nucleotide to correct a particular mutation. Moreover the Cas9 PAM is G-rich so it is not ideal for targeting organisms with AT-rich genomes, such as the malaria *Plasmodium falciparum*¹²⁵. For such applications, there is the option to use Cpf1, another RNA-guided endonuclease of the type II CRISPR system, with the “TTN” PAM sequence¹²⁶. It is also possible to alter the PAM specificity by mutating Cas9, as have been demonstrated by Kleinstiver and colleagues¹²⁷. Increasing the repertoire of Cas proteins with different PAM specificities will offer researchers greater flexibility in choosing target sites.

1.3. *Tcf7l2* as a candidate gene for study

1.3.1. Tcf/Lef proteins and Wnt signalling

Tcf7l2 is a member of the Tcf/Lef family of transcription factors that act downstream of the canonical the Wnt signalling pathway. The Tcf/Lef proteins act as transcriptional switches in regulating the transcription of Wnt target genes¹²⁸ (Figure 1.3.1A-B). In the absence of Wnt signalling, the destruction complex targets β -catenin for degradation by the proteasome pathway. The destruction complex comprises of several key proteins, namely APC, Axin and GSK3, CK1 and β -TrCP. Axin acts as a scaffold to hold components of the destruction complex and β -catenin together^{128, 129}. CK1 and GSK3 β sequentially phosphorylate β -catenin and create a docking site for the E3 ubiquitin ligase β -TrCP¹³⁰. This docking site is protected from de-phosphorylation by APC¹³¹, hence ensuring the ubiquitinylation of β -catenin. Ubiquitinylated β -catenin is then degraded by the proteasome¹³². Therefore Tcf/Lef proteins in the nucleus remain bound to the Groucho co-repressor and transcription is repressed. The availability of Wnt ligand is regulated by secreted sFRP and WIF, which bind directly to Wnt ligand and block its interaction with the Frizzled-LRP receptor complex^{133, 134}. Additionally, DKK can also interact with LRP¹³⁵ and cause it to be internalised¹³⁶, thus turning off Wnt signalling. In the presence of Wnt signalling, APC and Axin2 are sequestered to the Frizzled-LRP receptor complex, possibly via interactions with Dishevelled¹²⁸. Hence β -catenin is not degraded and instead accumulates in the cytoplasm. It eventually translocates into the nucleus and displaces Groucho from Tcf/Lef proteins. Co-activators and general transcription factors are recruited to the complex, thus activating transcription. The canonical Wnt pathway regulates cells proliferation and differentiation, and is especially important for the maintenance of adult stem cells¹³⁷.

Figure 1.3.1. The canonical Wnt signalling pathway.

(A) In the absence of Wnt signalling, β -catenin is ubiquitinated by the destruction complex and targeted to the proteasome for degradation. Tcf/Lef proteins in the nucleus remain bound to the Groucho co-repressor and transcription of Wnt associated genes is repressed. The availability of Wnt ligand is regulated by naturally secreted sFRP and WIF, which sequester Wnt ligand. **(B)** In the presence of Wnt signalling, components of the destruction complex are bound by the Frizzled-Lrp receptor and rendered inactive. Hence β -catenin is not degraded and instead translocates into the nucleus. There, it forms a complex with Tcf/Lef proteins and displaces Groucho. Co-activators are recruited to the complex and Wnt associated genes are transcribed. This figure is adapted from reference 128, Copyright © 2012 Elsevier Inc.

Within the canonical Wnt signalling pathway, negative feed-back and positive feed-forward loops are intricately interwoven to provide exquisite control over the strength of Wnt signalling activation. Components of the Wnt pathway such as Axin2, Lgr5 and many more, are Wnt target genes themselves. Axin2, which is part of the destruction complex, is a direct target of Tcf/Lef proteins. Its promoter region contains eight Tcf/Lef binding sites and abolishing these sites prevents Wnt-dependent upregulation of Axin2¹³⁸. Moreover β -catenin knock down in the SW620 colorectal cancer cell line caused a decrease in Axin2 levels¹³⁹. Thus Axin2 acts in a negative feed-back loop to regulate the level of Wnt signalling activation. Conversely, Lgr5 functions in a positive feed-forward loop to promote Wnt pathway activation. Lgr5 is a target gene of the Wnt signalling pathway because it is downregulated when the Wnt pathway is inhibited, by overexpressing dominant negative Tcf712, in colorectal cancer cells¹⁴⁰. Wnt pathway is negatively regulated by the action Rnf43 and Znf3, which ubiquitinylates Frizzled receptor and targets it for lysosomal degradation^{141, 142}. However in the presence of R-spondin, Lgr5 removes Rnf43 and Znf3 from the plasma membrane, hence halting degradation of Frizzled receptor¹⁴¹. The two examples provided here are not exhaustive, as there are many other feed-back or feed-forward loops in the Wnt signalling pathway. The existence of these loops alludes to the importance of balancing canonical Wnt pathway activation for homeostasis and yet avoiding a hyperproliferative state that could lead to cancer.

1.3.2. Wnt signalling and Tcf7l2 in intestinal homeostasis

The canonical Wnt signalling pathway is important for small intestine homeostasis. It has been shown that inhibiting the Wnt signalling pathway, through Dickkopf overexpression or β -catenin knock out, affected proliferation of epithelial cells and caused a complete loss of the crypts in the intestine of adult mice¹⁴³⁻¹⁴⁵. On the flip side, leaving the Wnt signalling pathway unchecked can also be detrimental. In mice with one mutant allele of the APC gene ($APC^{Min/+}$), spontaneous mutation of the wildtype allele leads to a loss of heterozygosity at the APC locus ($APC^{Min/-}$)¹⁴⁶. This causes intestinal epithelial cells to hyper-proliferate, forming polyps¹⁴⁷ that could develop into intestinal cancer. Wnt signalling promotes proliferation through multiple mechanisms. One of which is through expression of the Wnt target gene EphB that is important for organisation of the crypt structure¹⁴⁸, hence maintenance of the stem cell niche. EphB is expressed by cells in the crypt base, in a counter gradient to ephrinB ligand that is expressed by cells at the top of the crypt (Figure 1.3.2). When cells with the receptor and ligand meet, a repulsive force pushes the two apart¹⁴⁹. Therefore EphB expressing cells, especially paneth cells, are kept at the crypt base¹⁴⁸ where they contribute to maintenance of the stem cell niche.

Out of all the four Tcf/Lef proteins, Tcf7l2 is the most abundant member in the adult small intestine and colon¹⁵⁰, suggesting that it has an important function in these organs. It certainly plays an essential role in the small intestine during embryogenesis and foetal development, because germline Tcf7l2 knock out is perinatal lethal^{151, 152}. The mice die due to major disruption of the small intestine epithelial¹⁵¹. In order to study the role of Tcf7l2 in the adult intestine, an inducible CreER^{T2} Tcf7l2 knockout mouse was created^{153, 154}. Knock out was restricted to the intestine only because the Villin promoter was used to drive expression of CreER^{T2}^{153, 154}. By knocking out Tcf7l2 in adult mice, it was demonstrated that Tcf7l2 is important for epithelial proliferation and maintenance of Lgr5⁺ stem cells in the small intestine¹⁵³. This result fits with the expression profile of Tcf7l2 in the small intestine because Tcf7l2 is expressed through out the length of the crypt and villus¹⁵⁵ (Figure 1.3.2). It seems that Tcf7l2 is also important in the colon because knock out in the colon epithelium abolished epithelial stem and progenitor cell proliferation, as evidenced by the complete lack of Ki67⁺ staining in the crypt base¹⁵³ where the Lgr5⁺ stem cells are located¹⁴⁰. However the involvement of Tcf7l2 in epithelial stem and progenitor cell proliferation may be an indirect one. This is because Tcf7l2 is highly expressed by differentiated epithelial cells at the top of the crypts, but not in the stem cell compartment at the crypt base¹⁵⁵.

Therefore there is still controversy over the role of Tcf7l2 in the colon. It is difficult to say whether it is acting as a transcriptional activator or repressor. Resolving this may also provide insights into the role of Tcf7l2 in colorectal cancer.

Figure 1.3.2. The healthy epithelial of the small intestine and colon.

In the small intestine $Lgr5^+$ stem cells reside at the base of the crypt, interspaced with paneth cells. Paneth cell secrete Wnt ligands, which partly contributes to maintenance of the stem cell niche. $Lgr5^+$ stem cells actively proliferate and give rise to transit amplifying (TA) cells. As the TA cells move upwards into the villus, they differentiate and give rise to the different intestinal cell types. Cells at the end of their lifespan are shed at the top of the villus by anoikis. The +4 stem cells are typically non-dividing cells unless an injury is sustained. Intestinal epithelial cell turnover occurs every 3 to 5 days. In the colon, the $Lgr5^+$ stem cells reside at the base of the crypts but paneth cells are absent. Instead, the stem cell niche is entirely maintained by the mesenchyme. As $Lgr5^+$ stem cells proliferate, the TA cells move upwards and differentiate into the various epithelial cell types. Colon epithelial turn over occurs every 5 to 7 days. The expression data of Tcf/Lef proteins, determined by RNA *in situ* hybridisation, are indicated by purple bars. This figure is adapted with permission from reference 316 (see Appendix D for permission agreement). Information on the Tcf/Lef expression and signalling gradients are from references 155 and 317 respectively.

1.3.3. Is Tcf7l2 an oncogene or tumour suppressor in colorectal cancer?

Colorectal cancer is the third most common cancer worldwide¹⁵⁶. It is clinically graded into stages I to IV, with stage IV being the most severe because widespread metastasis has occurred. The best survival rates are observed with early intervention at stage I and II¹⁵⁷. Survival drops significantly once the cancer has metastasised to distant organs¹⁵⁷, such as the liver or lung^{158, 159}.

The progression of colorectal cancer occurs in a step-wise manner, through the accumulation of activating mutations in oncogenes and inactivating mutations in tumour suppressor genes. According to the Vogelgram model, driver genes are typically mutated in the following order during colorectal cancer progression: APC, Kras, p53 and lastly Smad4^{160, 161}. Mutations in the APC gene commonly occur in the mutation cluster region that spans amino acids 1282 to 1581^{162, 163}. This results in truncated APC proteins that are either diminished in their ability to bind β -catenin or lacking the domain involved in β -catenin destruction^{162, 163}. When both alleles of the APC gene are mutated, β -catenin is stabilised and the canonical Wnt signalling pathway is hyper-activated¹⁶⁴. Although this is the initiating event in colorectal cancer, mid to late stage colon tumours are also addicted to a high level of Wnt pathway activation. Advanced colon tumours created by mutating APC, Kras and p53, regressed when APC function was restored and Wnt signalling was brought back to a normal level¹⁶⁵. Targeting the Wnt signalling pathway is an attractive therapeutic opportunity because it seems to be indispensable during colorectal cancer progression.

Indeed, researcher and clinicians alike have recognised the potential of targeting the Wnt pathway for cancer therapy. Much of the focus has been on the components that act

upstream in the Wnt signalling pathway, such as Porcupine, Frizzled receptor and R-spondin ligand¹⁶⁶⁻¹⁶⁸. But as explained in Section 1.3.1, the canonical Wnt signalling pathway is self-regulated through complex feed-back and feed-forward loops. There is also extensive cross talk between the Wnt pathway and other signalling pathways such as the Hedgehog and Hippo pathways¹⁶⁹⁻¹⁷¹. This makes it difficult to fully predict the outcome of targeting any one component. Therefore it is important to explore the untapped potential of other Wnt signalling components.

Tcf7l2 is thought to be an oncogene that potentiates hyperactive Wnt signalling in colorectal cancer, because it is the most abundantly expressed Tcf/Lef protein in the colon. However there is no basis in this assumption. Instead, there are several lines of evidence to suggest that Tcf7l2 is a tumour suppressor. First, Tcf7l2 knock down in DLD-1 colorectal adenocarcinoma cell line resulted in increased proliferation¹⁷². Second, Tcf7l2 is expressed at lower levels in colon cancer biopsies compared to the matched healthy tissues¹⁵⁰. So far, there is only weak evidence to link Tcf7l2 to colorectal cancer. However it is a promising candidate gene for further investigation because it is at the heart of the Wnt signalling pathway in the colon.

1.3.4. Available transgenic mouse strains for studying Tcf7l2 gene function

Efforts to understand the role of Tcf7l2 in adult homeostasis and disease have been hampered by the lack of adequate transgenic mouse strains. Most of the available Tcf7l2 strains are germline or conditional knock outs (Table 1.3.4), hence fine control over *Tcf7l2* expression is not attainable with these strains. Germline knockout strains are unsuitable because they do not survive past the perinatal stage^{151, 152}. The cause of

death is due to disruption of the small intestine epithelial lining¹⁵¹. To overcome this, conditional ready strains were created by floxing either exon 1 or exon 11 of the *Tcf7l2* gene^{153, 173-175}. Tissue-specific knock out was achieved by crossing the floxed *Tcf7l2* strain to another strain that expresses the Cre recombinase in pancreas¹⁷⁴. This double transgenic mouse replicated many aspects of type 2 diabetes and was used to show that the loss of *Tcf7l2* in the pancreas resulted in decreased β cell function¹⁷⁴. However diabetes is a systemic disease that involves multiple organs, so a pancreas-specific knock down is not an ideal model. In another application, the floxed *Tcf7l2* strain was mated to a strain expressing tamoxifen-inducible CreER^{T2} recombinase^{153, 154}. *Tcf7l2* knock out was induced by tamoxifen treatment, but only in the gut because the intestine-specific Villin promoter was used to drive CreER^{T2} expression^{153, 154}. This tamoxifen-inducible mouse was used to investigate the role of *Tcf7l2* in small intestine and colon stem cell maintenance. However it needs to be adapted if it is to be applied to colorectal cancer studies, because Villin is down regulated in some colon tumours¹⁷⁶. The solution to this would be to mate the floxed *Tcf7l2* strain to a one that ubiquitously expresses CreER^{T2} for inducible but not conditional *Tcf7l2* knock out.

There are also two *Tcf7l2* overexpression strains. In the first strain, Krt14 promoter drives ectopic expression of *Tcf7l2* in keratinocytes¹⁷⁷. In the second strain, *Tcf7l2* is overexpressed by 2 to 3-fold in an expression pattern similar to the endogenous profile¹⁵². As pointed out in section 1.2.3, overexpression models are not ideal because of potential pleiotropic effects.

Looking at the collection of available *Tcf7l2* transgenic mouse strains, it is clear that none of them can provide the user with precise control over *Tcf7l2* expression. This level of control would be helpful when trying to understand *Tcf7l2* function in intestinal

Table 1.3.4. Available Tcf7l2 transgenic mouse strains.

Type	Location	Mutation Details	Reference
Knockout	Exon 10	5' end of exon 10 replaced by PGK-hygromycin selection cassette	151
Knockout	Exon 11	Double-stranded break repair results in frame-shift mutation and premature stop codon	152
Knockout	Exon 1	Translated part of exon 1 and splice donor replaced by Cre-ERT2 fusion	318
Knockout / Reporter tagged	Exon 1	Whole exon replaced by GFP-Cre fusion	173
Knockout / Reporter tagged	Exon 1	Translated part of exon 1 and splice donor replaced by GFP-Cre fusion	319
Conditional / Inducible	Exon 1	Floxed exon. Crossed with transgenic mice expressing Cre recombinase in a subset of Tcf7l2 expressing cells. Therefore Tcf7l2 is only knocked out in less than 1% of the cells.	173
Conditional / Inducible	Exon 1	Floxed exon. Crossed with transgenic mice expressing Cre recombinase under the control of the Pdx1 promoter, for pancreas-specific <i>Tcf7l2</i> knock out.	174
Conditional / Inducible	Exon 11	Floxed exon. Crossed with transgenic mice expressing Cre-ERT2 under the control of the Villin promoter. <i>Tcf7l2</i> knock out in the intestine epithelial is induced by tamoxifen.	153
Conditional / reporter tagged	Intron 4 / intron 5	L1L2_Bact_P cassette and LoxP inserted into intron 4 and intron 5 respectively (exon 5 is flanked by LoxP sites). Flp recombination restores gene activity and creates a conditional ready allele (subsequent Cre recombination deletes exon 5). Cre recombination alone creates a LacZ reporter knockout mouse.	175 and 320
Overexpression	Random	Krt14 promoter-Tcf7l2 expression cassette inserted. Krt14 drives expression in keratinocytes.	177
Overexpression	Random	Modified human Tcf7l2 BAC inserted. The BAC was modified by recombining full length mouse Tcf7l2 cDNA into the human Tcf7l2 translational start site.	152

homeostasis and colorectal cancer because it seems that Tcf7l2 has a very complex role in these processes, which could even be changing as the cancer progresses. Therefore an inducible and reversible Tcf7l2 transgenic mouse would be an excellent tool for clarifying Tcf7l2 function in colorectal cancer pathology.

1.4. Establishing a Platform for Novel Drug Target Discovery

1.4.1. The proposed platform

The key to successful drug discovery and development lies in choosing a robust target. Therefore it is important to uphold the most stringent criteria when validating potential drug targets. It is for this reason that I have decided on a genetics platform in mouse. With transgenic mice, it is possible to achieve precise control over gene expression and in a physiological setting. Out of all the mouse transgene systems that have been described, the Tet-inducible system is the most versatile because the target gene can be regulated in an inducible and reversible manner. This is an essential feature for the platform because the platform has to be applicable to a large variety of candidate genes. If spatial or temporal restrictions are imposed, it will limit the choice of candidate genes.

I propose a platform for inducible and reversible repression of a target gene (Figure 1.4.1). The platform is based on the Tet-inducible expression system and comprises of two transgenic mouse strains: (1) a strain that ubiquitously expresses the reverse tetracycline trans-silencer (rTetR-KRAB) in all tissues and (2) a strain where the tetracycline response element (TRE) has been inserted in an intron just downstream of

the candidate gene's promoter. The two transgenic strains will be inter-crossed to obtain double transgenic mice of appropriate genotypes. In the presence of doxycycline, a tetracycline analogue, rTetR-KRAB binds to TRE and represses transcription of the candidate gene. Repression is reversible on cessation of doxycycline treatment. To achieve partial repression, the doxycycline dose can be titrated or mice that are heterozygous for the TRE allele can be used.

The TRE strain is specific for the candidate gene. So a new TRE strain has to be created for every gene that the user is interested in studying. These can be paired with the generic rTetR-KRAB strain to obtain double transgenic mice with functional Tet-inducible repression system.

1.4.2. Aims and objectives

The aim of this work is to establish a platform for stringent drug target validation. The basis of this platform is an experimental mouse where the expression of a candidate gene can be regulated in an inducible and reversible manner (Figure 1.4.1). This requires the generation of two transgenic mouse strains. As a demonstration that the system is functional, I have chosen Tcf7l2 as the candidate gene because of its promising role in colorectal cancer. Once this platform has been established, I will shift my focus to investigate whether Tcf7l2 is a good therapeutic target for colorectal cancer. However I would like to point out that this platform has wider applications besides drug target validation. It can also be used to study the biological function and pathological role of other target genes. Since repression of the target gene is systemic, this platform is aptly suited for studying genes that are involved in multi-organ processes or disease. Therefore desirable candidates include metabolic, cancer and immune response genes.

Figure 1.4.1. The genetics platform for inducible and reversible gene regulation.

This platform is based on the Tet-off inducible expression system and requires two independently generated transgenic mouse strains. For the first strain, the tetracycline trans-silencer (rTetR-KRAB) expression cassette is inserted into the ROSA26 locus. Expression of rTetR-KRAB is ubiquitous since it is driven by the CAG hybrid promoter. For the second strain, the tetracycline response element (TRE) is inserted within an intron that is proximal to the candidate gene's promoter. The two transgenic strains will be mated to obtain the experimental mouse. In non-treated mice, the candidate gene is expressed as per normal because rTetR-KRAB has no effect alone. On administering doxycycline (Dox), rTetR-KRAB is activated and binds to TRE and represses the candidate gene. Repression is reversible on cessation of doxycycline treatment.

To achieve the aim, I have set out a sequential action plan:

- (1) Create the rTetR-KRAB trans-silencer and TRE mouse strains. This will be done using the indirect mouse transgenesis method, with CRISPR/Cas9 technology to enhance targeting efficiency;
- (2) Develop a suitable doxycycline treatment regimen. I will test various doxycycline doses and delivery routes to find the optimal treatment regimen;

Utilise the Tcf7l2 transgenic mouse strain to investigate the relevance of Tcf7l2 in intestinal homeostasis and cancer. It is anticipated that there will be a lag time between obtaining the founder mouse and establishing the double transgenic mouse colony. In the interim, I will conduct some preliminary studies using small intestine organoids derived from the first litter of double transgenic mice.

CHAPTER 2

Materials and Methods

2.1. Mouse Transgenesis

2.1.1. sgRNA design

Potential sgRNAs were identified using the “CHOPCHOP” web-based program^{178, 179}. The desired target regions for the rTetR-KRAB(Kid1) and TRE strains were the first intron of the ROSA26 locus and the third intron of the *Tcf7l2* locus, respectively. Additional specified parameters were *Mus musculus* (mm10/GRCm38) species, canonical “NGG” PAM and off-targets with up to 3 mismatches. The algorithm provides a list of possible sgRNAs with scores that are calculated based on the: (1) the number of off-target sites; (2) optimal GC content and; (3) a guanine at position 20^{178, 179}. All selected sgRNAs had a score less than 75 (low scores are more favourable), were 20 nt in length, began with a “G” at the 5’ end, located at least 150 bp from exon-intron boundaries and within 3 kb of the transcriptional start site (TSS).

2.1.2. Assembly of components for CRISPR/Cas9 gene targeting

For construction of the ROSA26-rTetR-KRAB(Kid1) replacement vector, the pL451 null vector backbone was used as a base to assemble the CAG promoter-rTetR-NLS-KRAB(Kid1)-poly(A) expression cassette. The rTetR sequence, without the VP16

sequence, was cloned from the rtTA vector¹⁸⁰. Partial Kid 1 sequence containing the minimal KRAB domain (amino acids 12 to 53)¹⁸¹ was derived by RT-PCR from mouse tissue. 5' and 3' homology arms corresponding to the ROSA26 gene locus were inserted to flank the expression cassette. This vector was constructed by my colleague, Dr Wu Zhihao. The vector map and sequence are provided in Appendix C. It should be noted that the KRAB domain used in Tet-Inducible trans-silencers are typically derived from Kox1. However in this thesis, the KRAB domain was derived from Kid1 protein. This was to avoid possible squelching effect, as explained in Section 4.1.2.

To facilitate insertion at various genomic loci, a basic vector with TRE was constructed. This was done by cloning the TRE_{mod} sequence from the pTREtight vector (Clontech) into the multiple cloning site of pL451 null vector backbone. The vector map and sequence of the pL451-TRE vector are provided in Appendix C. The TRE insertion locus can be defined by cloning different flanking 5' and 3' homology arms. To check the sequence of cloned homology arms, the universal T7 and T3 promoter primers can be used to prime the sequencing reaction.

To generate inserts for cloning of the Tcf7l2 homology arms, genomic DNA from W4 ES cells were amplified by PCR using KAPA Hifi DNA polymerase (Roche), according to the manufacturer's instructions. The primer sequences and PCR conditions are given in Appendix Table B5. The homology arms and the pL451-TRE vector were digested with the appropriate restriction enzymes, purified by gel extraction and ligated with T4 DNA ligase (New England Biolabs Inc.) by incubating for 10 min at room temperature. Ethanol precipitation was performed and the ligated vectors were introduced into electro-competent *E. coli* via electroporation at 1.8 kV in a 1 mm cuvette. The transformed bacteria were allowed to recover by shaking at 220 rpm for 1 h, in LB

broth. After recovery, transformed bacteria were plated onto LB agar plates, that contain 100 ug ml⁻¹ ampicillin, and grown overnight in a 37°C incubator. Vectors were extracted from transformed E. coli using MiniPrep™ Express Matrix (MP Biomedicals) according to the manufacturer's instructions. The vectors were sent to First Base Laboratories for verification by Sanger sequencing. All constructs used in subsequent experiments contained the cloned inserts in the correct orientation and were free of mutations. Large quantities of correctly cloned vectors were obtained by maxiprep using E.Z.N.A.® Plasmid Maxi Kit (Omega Bio-tek), according to the manufacturer's instructions.

The pX330-U6-Chimeric_BB-CBh-hSpCas9 vector was a gift from Feng Zhang (Addgene plasmid # 42230)¹¹⁶. This vector was modified by inserting the 2A-eGFP sequence between Cas9 and the bGH poly(A) signal, thus creating the pSpCas9(Chimeric_BB)-2A-GFP vector. sgRNA were cloned into the modified pX330-2A-GFP vector by following the method specified by the vector's creator¹¹⁶. Briefly, the sgRNA inserts were generated by annealing two oligos of 24nt in length. The annealed oligos had sticky ends that were complementary to Bbs1 restriction site, for directional cloning, hence they could be ligated directly to Bbs1 digested pX330 vector. The sgRNA sequences are provided in Appendix Table B1. The sgRNA insertion site was located between the U6 promoter and Chimeric_BB sequence.

2.1.3. Embryonic stem cell culture

W4 embryonic stem (ES) cells¹⁸² were grown on mouse embryonic fibroblast (MEF) feeder layer, in DMEM supplemented with 15% FBS, 1 X non-essential amino acids, 1

mM sodium pyruvate, 2 mM L-glutamine, 50 Units ml⁻¹ penicillin-streptomycin, 55 nM 2-mercaptoethanol and 1000 Units ml⁻¹ LIF, at 37°C and 7% CO₂. The medium was replaced daily and cells were passaged at 80% confluency with a 1:8 to 1:12 split ratio. Only cells that were below passage 30 were used for blastocyst microinjection.

For maintenance and expansion of MEF cells, the cells were grown on tissue culture coated dishes in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 50 Units ml⁻¹ penicillin-streptomycin and 55 nM 2-mercaptoethanol, at 37°C and 7% CO₂. The medium was replaced every two days and cells were passaged at 95% confluency with a 1:4 split ratio. To prepare feeder cell layer for culturing ES cells, MEF cells were grown to 100% confluency and irradiated with a dose of 30 Gy by exposure to a cesium-137 source. After irradiation, the medium was replaced and MEF cells were incubated for 1 h to 12 h, before the medium was changed to ES cell medium and ES cells were seeded. MEF cells used as feeder layer were not cultured for more than passage 7. Stocks of passage 0 MEF cells were isolated from mouse embryos and prepared by my colleague, Mr Arun Kumar.

MEF cells were cryopreserved in 90% FBS with 10% DMSO, and stored in liquid nitrogen. ES cells were cryopreserved in 50% ES cell medium with 40% FBS and 10% DMSO, and stored in liquid nitrogen. MEF and ES cells were frozen with a 1:4 and 1:6 split ratio, respectively. For reconstitution, the cells were rapidly heated in a 37°C water bath till thawed, and immediately diluted 1:15 in pre-warmed medium. MEF cells were plated directly onto tissue culture coated dishes while ES cells were washed once in medium before seeding onto MEF feeder layer.

2.1.4. Transient ES cell transfection

To prepare ES cells for transfection, both ES cells and the MEF feeder layer were detached from a 10 cm dish by using 0.05% trypsin-EDTA. MEF cells were depleted by re-plating onto the same tissue culture dish and incubating for 45min. Once MEF cells adhered to the dish, non-adhered ES cells were collected by vigorous pipetting, washed once with medium and resuspended in fresh ES cell medium at a 1:40 to 1:60 dilution.

Transient transfection of ES cells was achieved by using Fugene® HD (Promega Corporation), according to the manufacturer's recommendations. For validating sgRNAs, 2 µg of pSpCas9(Chimeric_BB)-2A-GFP vector was diluted in DMEM (without FBS), mixed with Fugene Transfection Reagent in a 1:5 ratio and incubated for 15 min at room temperature. An equal volume of ES cells was combined with the transfection mix and incubated for 2 h at 37°C. For producing targeted ES cell clones the same steps were performed, except 1 µg of pSpCas9(Chimeric_BB)-2A-GFP vector and 1 µg of pL451-TRE replacement vector were used. After transfection, ES cells were plated onto irradiated MEF feeder layer in 6 well plates. Transfected ES cells were harvested 72 h later for fluorescence activated cell sorting.

2.1.5. Fluorescence activated cell sorting

ES cells were harvested, depleted of MEF feeder cells, washed once in fresh medium and filtered through a 40 µm nylon mesh. Cell sorting was performed using a FACSAria™ cell sorter (BD Biosciences) with FACS Diva™ software (BD Biosciences). Non-transfected ES cells were used as the GFP⁻ control. Stringent gating

was performed to exclude doublets and the collection gate was set to encompass only cells that expressed high levels of GFP. For validating sgRNAs, GFP⁺ ES cells were pooled, cultured on MEF feeder layer and harvested for genomic DNA after 48 h. For producing targeted ES cell clones, single GFP⁺ ES cells were sorted into each well of a 96-well plate and cultured for five days until single colonies were visible. In both applications, sorted cells were maintained in medium containing 10 µg ml⁻¹ ciprofloxacin for at least three days before switching to the standard medium formulation without ciprofloxacin.

2.1.6. sgRNA validation

Genomic DNA from sorted ES cells was isolated by using the E.Z.N.A.® Tissue DNA kit (Omega Bio-tek Inc.), according to the manufacturer's instructions. PCR was performed using GoTaq® DNA Polymerase (Promega Corporation) to amplify the region spanning the sgRNA target site, and the products were analysed by agarose gel electrophoresis, stained with ethidium bromide and visualised using a 365 nm UV trans-illuminator. All PCR reactions yielded a single band of the expected size. The band was excised and purified with the E.Z.N.A.® Gel Extraction kit (Omega Bio-tek Inc.), according to the manufacturer's instructions. Purified PCR products and sequencing primers were sent to the third party company 1st BASE (First BASE Laboratories Sdn. Bhd.), for DNA sequencing by Sanger method. The sequence chromatograms were examined manually to determine whether cleavage took place.

2.1.7. ES cell clone screening

Genomic DNA from ES cell clones was isolated by using the E.Z.N.A.® Tissue DNA kit (Omega Bio-tek Inc.), according to the manufacturer's instructions. Screening PCR was performed using GoTaq® DNA Polymerase (Promega Corporation) or KAPA Hifi DNA polymerase (Roche), according to the manufacturer's instructions. Primer sequences and annealing temperatures are given in Appendix Table B2. PCR reaction products were analysed by agarose gel electrophoresis and stained with ethidium bromide. Visual capture was performed on a Gel Documentation system, fitted with a 365 nm trans-illuminator.

2.1.8. Blastocyst microinjection

All animal procedures were performed according to The National Advisory Committee For Laboratory Animal Research (NACLAR) guidelines and approved by the Institutional Animal Care and Use Committee (IACUC) at Nanyang Technological University (NTU).

Super-ovulated C57BL/6 female mice (4 to 5 weeks of age) were mated to C57BL/6 stud males, and pre-implantation blastocysts were harvested from the females at 3.5 days post-coitum by flushing the uterine horns with M2 medium (Sigma-Aldrich). ES cells were harvested, depleted of MEF feeder cells and 12 to 16 ES cells were injected into the blastocoel cavity of each blastocyst. Prior to implantation, injected blastocyst were collected into M16 medium (Sigma-Aldrich) and incubated at 37°C and 5% CO₂. About 8 to 12 injected blastocysts were implanted into each uterine horn of

pseudopregnant ICR females at 2.5 days post-coitum. Pseudopregnant females had been prepared by mating ICR females that were in estrus to vasectomised C57BL/6 males. About 12 h after setting up the mating pairs, the ICR females were checked for the presence of vaginal plugs to confirm that sterile mating had occurred. Implanted females were allowed to carry the pups to full-term.

2.1.9. Chimera breeding and colony establishment

Male chimeric founders (F0 generation) were mated to C57BL/6 female mice and agouti coloured pups (F1 generation) were screened for the transgene using PCR primer pairs and conditions as described in section 2.1.7. One transgene-positive heterozygous mouse (F1 generation) from each transgenic strain was used to establish that strain by backcrossing to C57BL/6 mice. The status of the backcross is currently generation F8 and F10 for the ROSA26-rTetR-KRAB(Kid1) and Tcf7l2-TRE strains respectively. Subsequent generations of mice after the F1 generation were genotyped using the multiplex PCR strategy described in section 2.1.10.

2.1.10. Mouse genotyping

For preparation of genomic DNA, mouse tail clips were incubated in SNET buffer pH 8.0 (20 mM Tris-Cl pH 8, 5 mM EDTA pH8, 400 mM NaCl, 1% SDS) with 200 µg ml⁻¹ proteinase K, at 55°C and continuous shaking till the tail clips dissolved. Fur and debris were removed by centrifugation and DNA was precipitated by adding an equal volume of ice-cold isopropanol. The DNA pellet was washed once in 75% ethanol, air-dried and

dissolved in TE buffer pH 8.0. Proteinase K was inactivated by incubating at 65°C for 20 min.

Multiplex PCR was performed using GoTaq® DNA Polymerase (Promega Corporation). Primer sequences and annealing temperatures are given in Appendix Table B3. PCR reaction products were analysed by agarose gel electrophoresis and stained with ethidium bromide. Visual capture was performed on a Gel Documentation system, fitted with a 365 nm trans-illuminator.

2.1.11. mRNA expression analysis

The various organs, as shown in Figure 3.1.5 and Figure 3.1.6, were harvested from F1 generation transgenic mice and homogenised directly in TRIzol® Reagent (Invitrogen) by pulsing for 2 min. For every 50 mg of tissue, 1 ml of Trizol was used. To separate the muscle layer from the epithelial layer of the small intestine and colon, both organs were harvested, cut open longitudinally and washed 10 times with ice-cold PBS pH 7.4 (without Mg^{2+} and Ca^{2+}). Small intestines were incubated on ice in 5 mM EDTA for 1 h with gentle rocking, while colons were incubated at 37°C in 5 mM EDTA for 20 min with gentle rocking. The epithelial layer was dislodged from the muscle layer with 40 hard flicking motions by swinging the whole arm downwards. To each epithelial cell fraction, 4 ml of TRIzol® Reagent was added and the cells were pipetted up and down vigorously till a homogenous solution was obtained. The muscle layer was homogenised in TRIzol® Reagent by pulsing for 2 min. Samples were stored at -80°C until processed for RNA extraction.

To extract RNA from tissues preserved in TRIzol[®] Reagent, phase separation was performed according to the manufacturer's instructions. The aqueous phase was then transferred to an RNAsimple Total RNA Kit spin column (TIANGEN Biotech Beijing Co., Ltd.) and RNA was purified with an additional on-column DNase I treatment, according manufacturer's instructions. RNA concentration and purity were measured using a Nanodrop[™] (Thermo Scientific[™]). Only samples with ratios of A260/280 greater than 1.95 and A260/230 between 2.00 to 2.35 were used in subsequent applications. Single-stranded cDNA was synthesised from 2 µg of purified RNA using M-MLV Reverse Transcriptase (Promega Croporation) with oligo dT primers, according to the manufacturer's instructions.

End-point reverse transcriptase PCR analysis was performed using GoTaq[®] DNA Polymerase (Promega Corporation), according to the manufacturer's instructions. The primers were designed to be gene-specific (Appendix Table B4) and confirmed to have only one hit (to the mRNA that they were designed to) in a BLAST search against the "reference mRNA sequences" database and under the default settings. The reaction was performed using cDNA template for 40 cycles, and the products were analysed by agarose gel electrophoresis, stained with ethidium bromide and photographed on a Gel Documentation system, fitted with a 365 nm trans-illuminator.

Semi-quantitative reverse transcriptase PCR analysis was performed using GoTaq[®] DNA Polymerase (Promega Corporation), according to the manufacturer's instructions. The primers were designed to be gene-specific (Appendix Table B4) and confirmed to have only one hit (to the mRNA that they were designed to) in a BLAST search against the "reference mRNA sequences" database and under the default settings. The amount of cDNA template and all PCR reaction reagents were kept constant while the number

of cycles was varied. PCR was performed for 22, 25, 28, 31 and 34 cycles and the products were analysed by agarose gel electrophoresis, stained with ethidium bromide and photographed on a Gel Documentation system, fitted with a 365 nm trans-illuminator.

2.2. Development of A Doxycycline Treatment Regime

2.2.1. Immortalisation of bone marrow cells

The process to immortalise bone marrow progenitors consisted of multiple simultaneous steps. For a clearer picture, the process is presented as a timeline (Appendix Figure 2.2.1). The details for each step are provided below.

The pMYS-Nup98HoxB4-IRES-Puro retroviral vector was constructed previously by a former lab member. It was created by modifying the commercially available pMYS-IRES-GFP Retroviral Expression Vector (Cell Biolabs, Inc.). A map and sequence of this vector are provided in Appendix D.

The Platinum-E Retroviral Packaging Cell line (Cell Biolabs, Inc.) was used to package the retroviral vector. For routine maintenance, Platinum-E cells were cultured in IMDM supplemented with 10% FBS. The medium was replaced every alternate day and cells were passaged at 80% confluency with a 1:4 split ratio using 0.05% trypsin-EDTA.

To prepare for calcium chloride transfection, Platinum-E cells were seeded at a density of 30% the day before transfection. Cells were 70% confluent at the time of

transfection. Calcium chloride transfection was performed by combining 20 µg retroviral vector (pMYS-Nup98HoxB4-IRES-Puro), 124 mM calcium chloride and HBS pH 7.05 (274 mM NaCl, 10 mM KCl, 1.4 mM Na₂HPO₄, 15 mM D-glucose, 42 mM HEPES, pH adjusted with NaOH), with vigorous pipetting for aeration. The transfection complexes were added to Platinum-E cells with fresh medium containing 25 mM chloroquine and incubated for 12 hours. After incubation, chloroquine was removed by changing to fresh medium. On day 3, the medium containing viral particles was harvested and replaced with fresh medium.

To prepare for spin infection, 6-well tissue culture plates were coated with 20 µg ml⁻¹ retronectin overnight at 4°C, on days 2 and 3. On the day of spin infection, supernatant containing viral particles and 5 µg ml⁻¹ polybrene was added to the retronectin-coated plates, and the plates were centrifuged at 3,000 x g for 2 h at 4°C to adhere viral particles.

To prepare for bone marrow harvest, female F1 generation ROSA26-rTetR-KRAB(Kid1) wildtype and heterozygous mice (4 weeks of age) were administered 5-fluorouracil (160 µg g⁻¹ body weight) by intra-peritoneal injection. On day 3, bone marrow cells were flushed from the femur and tibia bones with IMDM containing 2% FBS, depleted of red blood cells using 0.89% ammonium chloride solution. Cells were seeded at a density of 2.5 million cells in each well of the 6 well plate containing adhered particles. Spin infection was performed by centrifuging at 580 x g for 1 h at room temperature. Following infection cells were incubated at 37°C with 5% CO₂ for 1 h, and the medium was changed to IMDM supplemented with 2% FBS, IL-6 conditioned medium and SCF conditioned medium. Both IL-6 and SCF were homemade by the lab and were previously titrated to determine optimum dilution factor.

On day 4, spin infection was performed a second time for bone marrow cells that were infected the day before. The procedures were performed as done on day 3. Bone marrow cells were initially passaged at a 1:4 split ratio when they reached 80% confluency, but the split ratio was increased to 1:6 or 1:8 after several passages. Over time, immortalised cells overtook the culture while non-immortalised cells differentiated and stopped proliferating. By 1.5 weeks in culture, most of the remaining cells were immortalised cells. Therefore it was not necessary to perform antibiotic selection for immortalised cells.

2.2.2. Production of GFP reporter bone marrow cell line

To create GFP reporter bone marrow lines that are responsive to rTetR-KRAB(Kid1) regulation, the pMYS-Nup98HoxB4-IRES-Puro retroviral vector was modified to create the pMYS-TRE-IRES-GFP retroviral vector. This was done by replacing Nup98HoxB4 and Puromycin resistance gene with the TRE sequence and GFP respectively. A map and sequence of this vector are provided in Appendix D.

The TRE-IRES-GFP expression construct was stably integrated into immortalised bone marrow cells by retroviral transduction, according to the method described in Section 2.2.1. Transduced cells were expanded in culture, harvested, washed once in fresh medium and filtered through a 40 µm nylon mesh. Fluorescence activated cell sorting was performed using a FACS Aria™ cell sorter (BD Biosciences) with FACS Diva™ software (BD Biosciences). GFP⁻ and GFP⁺ cells were simultaneously sorted into separate tubes, and cultured in medium containing 10 µg ml⁻¹ ciprofloxacin for at least

three days before switching to the standard medium formulation without ciprofloxacin. Sorted cells were expanded in culture and checked for purity using a FACSCalibur™ (BD Biosciences) with CellQuest™ Pro software (BD Biosciences). All sorted cell populations were determined to be at least 94% pure.

GFP⁻ and GFP⁺ reporter bone marrow cells were cryopreserved in 90% bone marrow medium with 10% DMSO at a 1:6 split ratio, and stored in liquid nitrogen. The bone marrow cell lines were cultured for not more than 3.5 weeks after they were harvested from mice. For reconstitution, the cells were rapidly heated in a 37°C water bath till thawed, immediately diluted 1:10 in pre-warmed medium and washed once before plating in tissue culture coated flasks.

2.2.3. *In vitro* doxycycline kinetics assay

Doxycycline hyclate (Sigma) was dissolved in water to a concentration of 50 mg ml⁻¹ and diluted fresh in bone marrow medium to the appropriate concentrations. Reporter bone marrow cells were cultured in doxycycline medium for a period of 10 days. To determine induction kinetics, a cell aliquot was taken each day and the mean GFP fluorescence intensity was measured using a FACSCalibur™ (BD Biosciences) with CellQuest™ Pro software (BD Biosciences). The GFP intensities of heterozygous cells were normalised twice: first to the untreated control (no doxycycline) and second to wildtype cells cultured at the same doxycycline concentration.

2.2.4. Generation of bone marrow chimera mice

Recipient C57BL/6 mice (5 to 8 weeks of age) were irradiated twice with 5.5 Gy by exposure to a cesium-137 source (each dose given 4 h apart) the day before injection of bone marrow cells. Immortalised reporter bone marrow cells (total culture time of not more than 5 weeks) were washed three times to remove FBS and filtered through a 40 μ M nylon mesh. Recipient mice were anaesthetised with isoflurane and about 1 to 2 million cells were injected by retro-orbital route. Mice were given 2 mg ml⁻¹ neomycin in the drinking water for 1.5 weeks after irradiation, and were maintained for 7 to 9 weeks to allow hematopoietic reconstitution before beginning doxycycline treatment.

2.2.5. Doxycycline administration *in vivo*

Doxycycline was administered to mice either by intra-peritoneal injection or orally. For the intra-peritoneal route, doxycycline hyclate (Sigma) was dissolved in PBS to a concentration of 5 mg ml⁻¹ and administered at a dose of either 25 μ g g⁻¹ or 50 μ g g⁻¹ of body weight. For the oral route, mice were fed food pellets containing 0.2 g kg⁻¹, 0.625 g kg⁻¹ or 1 g kg⁻¹ doxycycline *ad libitum*. The food pellets were formulated according to the Teklad base diet 2018 (Envigo), which is a standard rodent maintenance diet.

2.2.6. Determination of GFP repression in bone marrow chimeras

To track GFP repression after doxycycline administration, blood was obtained from isoflurane-anaesthetised mice by retro-orbital bleeding using glass pipettes that were pre-coated with 2.5% sodium citrate solution. Drawn blood was immediately diluted about

1:5 with 2.5% sodium citrate solution. PBMCs were isolated from blood by Ficoll-Paque™ PLUS gradient (GE Healthcare Life Sciences) according to the manufacturer's instruction and washed twice in PBS. Fc block was performed and the cells were stained with an antibody cocktail containing Buv395-conjugated anti-CD19 (clone ID3, BD Biosciences), APC-conjugated anti-CD3 (clone 145-2C11, Biolegend), APC-Cy7-conjugated anti-CD11b (clone M1/70, Biolegend) and PE-conjugated anti-Gr-1 (clone RB6-8C5, Biolegend).

Spleen cells were harvested from sacrificed mice, by cutting the spleen into fine pieces in IMDM containing 2% FBS, digesting with 1 mg ml⁻¹ Collagenase D (Roche) at 37°C for 1 h and mashing to release the cells. Cells were prepared for staining by lysing red blood cells with 0.89% ammonium chloride, washing once with fresh medium and filtering through a 100 µm nylon mesh. Fc block was performed and the cells were stained with an antibody cocktail containing PerCp-Cy5.5-conjugated anti-Ly6c (clone Hk1.4, eBioscience), eFlour 450-conjugated anti-MHC II (clone M5/114.15.2, Invitrogen), Buv395-conjugated anti-CD19 (clone ID3, BD Biosciences), APC-conjugated anti-Ly6G (clone IA8, Biolegend), APC-Cy7-conjugated anti-CD11b (clone M1/70, Biolegend) and PE-Cy7-conjugated anti-CD11c (clone N418, eBioscience) and Buv737-conjugated anti-CD3 (clone 145-2C11, BD biosciences).

Acquisition of stained blood and spleen cells was performed using a BD LSRFortessa™ (BD Biosciences) with FACS Diva™ software (BD Biosciences). Data were analysed with FlowJo® X (Flowjo, LLC).

2.2.7. *Tcf7l2* expression analysis in doxycycline-treated mice

RNA isolation and cDNA synthesis were performed as described in Section 2.1.11. The relative cDNA concentrations were measured by quantitative PCR using PrecisionFAST qPCR Master Mix with SYBR green (Primerdesign Ltd), according to the manufacturer's instructions. The primers were designed to be gene-specific (Appendix Table B4) and confirmed to have only one hit (to the mRNA that they were designed to) in a BLAST search against the "reference mRNA sequences" database and under the default settings. Reactions were performed on 48-well plates in a total volume of 20 µl in duplicate. Primers were validated and determined to amplify only one product and had a two-fold amplification efficiency. Thermocycling parameters were as recommended by the manufacturer, with an annealing temperature of 60 °C and an extension time of 30 s. Real-time fluorescence was recorded with the Eco Real-Time PCR System (Illumina) and the threshold was calculated automatically by the software. Amplification specificity was verified by melting curve analysis (single peak) and separation of the qPCR products on an agarose gel (single band of the expected size). Relative expressions were calculated using the comparative C_T method ($2^{-\Delta\Delta C_T}$) by normalising to β -actin expression.

2.3. Investigation of *Tcf7l2* function in small intestine organoids

2.3.1. R-spondin1 production and purification

Recombinant mouse R-spondin1 IgG2a Fc chimeric protein was overexpressed from SCC111 cells (Merck Millipore) in DMEM supplemented with 2 mM L-glutamine,

according to the manufacturer's instructions. The supernatant was cleared of cells and debris by centrifugation at 3,000 x g for 20 min at 4°C, and filtered through a 0.22 µm vacuum filter. Protein was affinity purified by applying the supernatant to a gravity-flow column containing Protein A Agarose beads (Sino Biological) at 4°C. After binding (R-spondin1 captured by Protein A via Fc region), the column was washed with at least 10 column volumes of 100 mM NaCl and until the absorbance at 280 nm was at background levels. R-spondin1 protein was eluted in 50 mM glycine-HCl with 80 mM NaCl pH 2.8 and immediately neutralised with 2 M K₂HPO₄. Eluted protein was visualised on a denaturing SDS-PAGE gel (8% resolving and 5% stacking) stained with Coomassie blue dye (Appendix Figure A2.3.1), and concentration was determined with a UV spectrophotometer using a quartz cuvette with each unit of absorbance at 280 nm and 1 cm path length taken to be 1 mg ml⁻¹. Protein was diluted to 100 µg ml⁻¹ in elution buffer, snap frozen in liquid nitrogen and stored as single use aliquots at -80°C.

2.3.2. Organoid derivation

To establish organoid lines from double transgenic or ApcMin mice, the small intestines were harvested, cut open longitudinally and washed 10 times with ice-cold PBS pH 7.4 (without Mg²⁺ and Ca²⁺). This and all subsequent steps were performed quickly and on ice to ensure the viability of the epithelial stem cells. Intestines were incubated in 5 mM EDTA for 1 h with gentle rocking, and crypts were dislodged with 10 hard flicking motions by swinging the whole arm downwards. Four consecutive crypt fractions were collected, visualised under the stereo microscope, and the two fractions with the most number of intact crypts were pooled and filtered through a 70 µm cell strainer to remove villi. About 500 crypts were embedded in each 50 µl dome of Matrigel® Growth Factor

Reduced Matrix (Corning, #356231) on 48-well plates, and covered with Advanced DMEM/F12 supplemented with 50 ng ml⁻¹ recombinant mouse EGF (Invitrogen, #PMG8043), 100 ng ml⁻¹ recombinant mouse Noggin (PeproTech, #250-38), 1 µg ml⁻¹ recombinant mouse R-spondin1, 2 mM glutamax, 1 X B27, 1 X N2, 1 mM N-acetylcysteine, 10 mM HEPES, 50 U ml⁻¹ penicillin-streptomycin. The crypts were maintained at 37°C and 5% CO₂, and formed mature organoids after 5 to 6 days in culture.

2.3.3. Organoid maintenance

After derivation, organoids were maintained as 50 µl Matrigel domes covered with organoid medium, and passaged every 5 to 7 days. To passage, organoids were released from Matrigel domes by treating with Gentle Cell Dissociation Reagent (STEMCELL Technologies Inc.) according to the manufacturer's instructions, washed twice in Advanced DMEM/F12, re-embedded in Matrigel at a 1:3 or 1:4 split ratio and covered with organoid medium.

Organoids were cryopreserved in Bambanker™ (Wako Pure Chemical Industries Ltd.) at 1:1 split ratio, and stored in liquid nitrogen. For reconstitution, the organoids were rapidly heated in a 37°C water bath till thawed, pelleted by centrifugation at 200 X g for 5 min, embedded in Matrigel and cultured in organoid medium supplemented with 10 µM Y-27632. The medium was switched back to the standard medium formulation without Y-27632 after 4 days in culture.

2.3.4. Assaying function of Tcf7l2 in organoids

Tcf7l2 was repressed in small intestine organoid lines by changing to fresh organoid medium containing 1 $\mu\text{g ml}^{-1}$ doxycycline. Doxycycline medium was refreshed every alternate day. Photographic capture was performed with a CCD camera fitted onto a light microscope. For assessment of budding, the organoids were scored manually into four categories depending on the number of buds. At least 30 organoids were scored for each treatment group. For assessment of organoid size, the 2D-area of each organoid was measured by tracing around the organoid, using ImageJ version 1.51s software (National institutes of Health, USA)¹⁸³. At least 30 organoids were measured for each treatment group. Statistical significance was determined using two-way ANOVA with Bonferroni post-hoc test, from two independent experiments.

CHAPTER 3

Results

3.1. Creating Transgenic Mouse Strains

3.1.1. General plan for creating transgenic mouse strains

The proposed Tet-inducible repression system requires two transgenic mouse strains. First, a general strain that ubiquitously expresses the rTetR-KRAB(Kid1) trans-silencer. Second, a gene specific strain where the target of the trans-silencer (TRE sequence) is in close proximity to the transcription start site of the gene of interest. When the strains are inter-crossed to get double transgenic mice, the Tet-inducible repression system becomes functional.

The first step in establishing this system is to create both mouse strains and I did this with assistance from Dr. Piotr Tetlak. The transgenesis method is outlined below. Details for each step are presented in subsequent sections 3.1.2 to 3.1.6.

The standard ES cell transgenesis technique pioneered by Capecchi, Evans and Smithies⁹⁸⁻¹⁰¹ was used to create both mouse strains. Essentially ES cells are targeted, via homologous recombination, so that they carry the desired gene modification. These modified ES cells are then introduced into blastocysts to generate chimeras. This process is depicted as a schematic for easy reference (Figure 3.1.1).

Figure 3.1.1. Approach for creating transgenic mouse strains.

Transgenesis method to generate new strains of mice in three steps. In the first step, potential sgRNAs are selected, cloned into the sgRNA-Cas9 expression vector and tested *in vivo* for their ability to guide Cas9 to the specific target locus for double-stranded DNA cleavage. Once a suitable sgRNA is identified, the replacement vector is constructed by cloning homology arms to flank the insert sequence. In the second step, the desired transgene modification is created in embryonic stem (ES) cells using CRISPR/Cas9 gene targeting technology. ES cell targeting is performed by transfecting the cells with both the sgRNA-Cas9 and replacement vectors. The Cas9-sgRNA complex creates double-stranded break at the target locus, which can then be repaired by homologous recombination using the replacement vector as a donor template. As a result of this, the desired sequence is inserted at the specified locus in targeted ES cells. In the third step, chimeras are generated using the targeted ES cells. Following microinjection of ES cells into host blastocysts, the injected blastocysts are implanted into pseudopregnanant mothers and carried to full term to give chimeras. Chimeras are mated to test for germline transmission. Pups that carry the desired gene modification are used to establish the strain.

ES cell targeting was performed with the help of CRISPR/Cas9 gene editing technology since it can deliver high targeting efficiencies; yet it is relatively easy to use compared to other nuclease-assisted methods like TALEN and ZFN¹⁰⁴⁻¹⁰⁶. To prepare for CRISPR/Cas9 targeting, sgRNAs were selected and tested to determine whether they can induce double-stranded DNA (dsDNA) break at the target locus. Following the identification of a suitable sgRNA, the replacement vector was constructed. Gene targeting was performed in ES cells, with the sgRNA-Cas9 vector and replacement vector, to obtain clones with the desired modification.

To generate chimeras, targeted ES cells were injected into blastocysts. Injected blastocysts were then implanted into pseudo-pregnant females and carried to full term. The resulting chimeras were bred to test for germline transmission. This was done by screening their progeny for inheritance of the gene modification. Positive progeny mice were used to establish the strain.

3.1.2. Identifying sgRNAs that are associated with specific nuclease activity

Successful CRISPR/Cas9 gene targeting relies on an sgRNA that is associated with specific nuclease activity. This means that the sgRNA must be able to guide Cas9 to the target site for induction of dsDNA break. An *in vivo* approach was adopted to test potential sgRNAs (Figure 3.1.2A). First, several sgRNAs with adjacent canonical PAM site (5'-NGG-3') were identified using the "CHOPCHOP" web program^{178, 179}. Next, individual sgRNAs were cloned into a modified pSpCas9(Chimeric_BB)-2A-GFP

vector. I had modified the original pSpCas9(Chimeric_BB) vector that was created by Cong and colleagues¹¹⁶, by cloning the 2A-GFP sequence downstream of Cas9. This was done to facilitate isolation of Cas9-expressing cells in a latter step. However the base vector already contains the sgRNA scaffold and human SpCas9, expressed from the U6 and CBh promoters respectively. After cloning in the variable sgRNA sequence to complete the sgRNA scaffold in the pSpCas9(Chimeric_BB)-2A-GFP vector, the vector was transfected into W4 ES cells. This is the same ES cell line that will be used for gene targeting and chimera generation. Finally GFP⁺ cells were isolated by fluorescence-activated cell sorting (FACS) at three days after transfection. dsDNA breaks generated by CRISPR/Cas9 are repaired with a half-life of about 1 to 10 hours¹⁸⁴. Hence a buffer period of three days should be sufficient for expression of the CRISPR/Cas9 components, DNA cleavage and repair to occur. Cas9-sgRNA-mediated dsDNA breaks are typically repaired by the error-prone non-homologous end joining (NHEJ) pathway, thus creating indel mutations at the target site¹⁸⁴.

Popular choices to detect such indel mutants include the Surveyor and T7 endonuclease assays. However I did not have the reagents on-hand, so Sanger sequencing was used instead. After dsDNA break and repair, cells in the bulk population differ from one another at the target locus because any number of bases can be inserted or deleted during repair. This population variability shows up as a multi-peaked chromatogram at the region surrounding the cleavage site. This is because a cell that undergoes an insertion or deletion mutation will have a sequence that is frame-shifted to the right or left respectively. When numerous indel sequences accumulate (from different cells), each with different frame-shift, the read-out becomes multi-peaked. In a scenario where there is no indel mutation, the chromatogram will be single-peaked.

Figure 3.1.2. Identifying sgRNAs that are associated with specific nuclease activity.

(A) *In vivo* assay to screen for sgRNAs that are associated with specific nuclease activity. Potential sgRNAs are selected, cloned into the modified pSpCas9(Chimeric_BB)-2A-GFP vector and transfected into embryonic stem (ES) cells. Successfully transfected cells express GFP and are isolated by fluorescence-activated cell sorting (FACS). In these cells, the Cas9-sgRNA complex creates a double-stranded break at the target locus that is then repaired by the non-homologous end joining (NHEJ) mechanism. Such events create indel mutations at the break site, which can be detected by PCR amplification across the target site and sequencing the PCR product. (B) The Sanger sequencing method, for detecting specific nuclease activity, is put to the test. The sequence chromatograms for five Tcf7l2-specific sgRNAs (T1 to T5) are lined-up, with their corresponding wildtype sequence. The positions of PAM, sgRNA and cleavage site indicated are in relative orientation to the sequenced strand. For T5, the PAM sequence

reads as 5'-“NGG”-3' since the sgRNA and PAM are in the same orientation as the sequenced strand. For T1 to T4, the PAM reads as 5'-“CCN”-3' because the sgRNA and PAM are orientated on the complementary strand. A short length of the upstream sequence is included to show that the sequence is clean—baseline peaks are absent or minimal. With the exception of sg 3, all the other sgRNAs had a multi-peaked sequence in vicinity of the cleavage site. Abbreviations (in alphabetical order): PAM protospacer adjacent motif; sgRNA single guide RNA; WT wildtype.

To demonstrate the utility of the sequencing method, five sgRNAs that target the *Tcf7l2* locus were examined (Figure 3.1.2C). Four of the sgRNAs displayed a multi-peaked chromatogram around the cleavage site, which would be expected in the event of dsDNA break and NHEJ repair. Only the chromatogram for sgRNA T3 was single-peaked suggesting that cleavage did not occur. Out of the four sgRNAs that are associated with obvious nuclease activity, sgRNA T1 was selected for subsequent ES cell targeting at the *Tcf7l2* locus.

For constitutive expression of rTetR-KRAB(Kid1) in the trans-silencer mouse strain, the rTetR-KRAB(Kid1) expression cassette was targeted to the transcriptionally permissive ROSA26 locus^{185, 186}. It was not possible to target the exact provirus integration site described in the original papers, due to the “NGG” PAM requirement of Cas9^{114, 187}. Instead, an sgRNA that was near (280 bp upstream) to the provirus integration site was selected. I decided to use the selected sgRNA (R1) directly for gene targeting in ES cells without prior efficiency testing, since the “CHOPCHOP”

program^{178, 179} had predicted that it would facilitate efficient DNA cleavage based on its sequence features¹⁸⁸.

Since ROSA26 and *Tcf7l2* target sites were determined, the relevant replacement vectors were constructed by cloning homology arms to flank the sequence that is to be inserted. The insert was the rTetR-KARB(Kid1) expression cassette for the ROSA26 locus and the TRE sequence for the *Tcf7l2* locus. It was only possible to do this after sgRNA testing because the homology arms had to correspond to the region surrounding the target site.

3.1.3. CRISPR/CAS9-mediated gene targeting in embryonic stem cells

With all the vector components ready, ES cell targeting could proceed. Figure 3.1.3A shows the timeline of an ES cell targeting experiment. W4 ES cells were transfected with the pSpCas9(Chimeric_BB)-2A-GFP vector (with the sgRNA sequence cloned in) and replacement vector. All the components required for CRISPR/Cas9-mediated dsDNA break at the target locus, namely the sgRNA and Cas9, were encoded by the pSPCas9(Chimeric_BB)-2A-GFP vector. The replacement vector contained either the rTetR-KRAB(Kid1) expression cassette or the TRE sequence for insertion into the ROSA26 or *Tcf7l2* gene locus, respectively. This ES cell line was chosen for its robust germline competency¹⁸². At three days post-transfection, cells that expressed high levels of Cas9 (GFP^{high} cells) were isolated by FACS. Stringent gating for single cells was performed to remove doublets (Figure 3.1.3B). The percentage of GFP^{high} cells ranged from 1.5% to 4%, across different experiments.

Figure 3.1.3. Targeting in embryonic stem cells using CRISPR/Cas9 technology.

(A) Timeline for embryonic stem (ES) cell targeting using CRISPR/Cas9 technology. Mouse ES cells (W4 cell line) are co-transfected with the pSpCas9(Chimeric_BB)-2A-GFP vector and replacement vectors. The pSpCas9(Chimeric_BB)-2A-GFP vector provides Cas9 and the sgRNA, in order to create a dsDNA break at the target locus. Thereafter, the replacement vector is required to insert the rTetR-KRAB(Kid1) expression cassette into the ROSA26 locus to get the ROSA-rTetR-KRAB(Kid1) mouse strain; or the TRE sequence into the Tcf7l2 gene locus to get the Tcf7l2-TRE mouse strain. Both Cas9 and GFP are co-expressed from the pX330-2A-GFP vector. Therefore cells that express high levels of Cas9 were isolated by sorting GFP^{high} cells using fluorescence-activated cell sorting (FACS), at three day post-transfection. To ensure clonality, single cells are sorted into individual plate wells. By day seven, the single cells form colonies. Wells that contain only single colonies were re-plated onto replica plates. DNA was harvested from one of the replica plates to screen for targeted gene modification. Positive clones are expanded from the other replica plate and either injected into blastocysts immediately or cryopreserved. **(B)** Gating strategy for isolating GFP-expressing ES cells by FACS. First, cells are identified in a forward/side scatter plot (FSC-A/SSC-A). Second, doublets are excluded using SSC-W/SSC-H and FSC-W/FSC-H plots. Finally, cells that express high levels of GFP (GFP^{high}) are sorted. Percentages of cells encompassed within the gates are given as pink-coloured numbers.

After sorting single cells into individual wells on a plate, the cells were allowed to grow until they formed microscopically visible colonies on day 7. At this point, wells with single colonies were identified and those with multiple colonies were discarded. This was crucial to preserve clonality. Single colony discrimination would have been difficult after this time point because colonies became so large that they overlap. Following this, clones were recovered and re-plated onto replica plates. DNA was harvested from one of the replica plates for screening by PCR. Positive clones, that had the desired modification, were expanded from the other replica plate and either injected into blastocysts immediately or cryopreserved for injection at a later date.

The results for the ES cell targeting experiments are summarised in Table 3.1.3. For ROSA26-rTetR-KRAB(Kid1) targeting, the recovery rate after single cell sorting was 6.3%. Out of the recovered clones, 8.3% (two clones) were correctly targeted. For Tcf7l2-TRE targeting, the recovery rate after single cell sorting was 10.6%. Only 24 of the recovered clones were screened, because targeting was so efficient that there was no shortage of correctly targeted clones for blastocyst microinjection. The targeting efficiency was 25% (six clones). For details of the targeting and screening strategies pertaining to each gene, please refer to sections 3.1.5 and 3.1.6.

3.1.4. Generating germline-transmitting chimeras from targeted embryonic stem cells

Following established practices, blastocysts were obtained from super-ovulated C57BL/6 female mice, injected with targeted ES cells (Figure 3.1.4A–F) and implanted into pseudo-pregnant mothers to obtain chimeras. The whole process was performed in-house.

Table 3.1.3. Summary of frequencies obtained at each stage of embryonic stem (ES) cell targeting process.

#	Target	sgRNA	Sorted cells	Recovered clones (% of sorted)	Screened clones	Correctly targeted clones (% of screened)
1	ROSA26-rTetR-KRAB(Kid1)	R1	384	24 (6.3%)	24	2 (8.3%)
2	Tcf7l2-TRE	T1	480	51 (10.6%)	24*	6 (25.0%)

*** Out of the 51 clones that were recovered, only 24 of them were screened to check for desired gene modification.**

Chimeras were obtained from both ROSA26-rTetR-KRAB(Kid1) and Tcf7l2-TRE targeted ES cell clones. They were tested for germline transmission by mating to C57BL/6 mice. If ES cells had contributed to the germ cells, the chimera's offspring would inherit the gene modification. Only male chimeras with predominantly agouti-coloured fur coat (Figure 3.1.4G) were mated because their gender and coat colour implied a high contribution from injected ES cells. The W4 ES cell line is karyotypically male and derived from the agouti coloured 129S6 mouse strain¹⁸². Mid to low chimerism mice that are mostly black in colour (Figure 3.1.4H), had a lower chance of transmitting through the germline.

F1 generation pups from chimera mating were identified by coat colour. The agouti allele is dominant to the black allele. So agouti-coloured pups originated from ES cells while black-coloured pups were from host blastocyst. Hence agouti pups were screened to check inheritance of the gene modification. F1 generation mice with the targeted ROSA26-rTetR-KRAB(Kid1) and Tcf7l2-TRE alleles were obtained and bred to establish the mouse strains. In the next two sections (section 3.1.5 and 3.1.6), the targeting strategies and screening methods for both mouse strains are given in detail.

Figure 3.1.4. Blastocyst microinjection, chimeras and germline transmission.

(A–F) Sequential photographs of a blastocyst being injected with targeted embryonic stem (ES) cells for chimera generation. (A) About 12 to 16 ES cells are collected in shaft of the injection pipette. (B) A blastocyst is positioned and held in place with a holding pipette. (C) The injection pipette, containing ES cells, is pushed through the blastocyst's zona pellucida and trophoblast and into the cavity. (D) Cells are expelled from the injection pipette and into the cavity. (E) Once all the cells are expelled, the injection pipette is withdrawn from the blastocyst. (F) The injection is now complete and the blastocyst is released. Injected blastocysts are implanted pseudopregnant females and chimeric pups are carried to full term. (G–H) Chimeras from blastocyst microinjection. The chimeras are made up of cells from two different origins. Contribution from ES cells (129S6 strain) and host blastocyst (C57BL/6 strain) are marked by agouti and black fur respectively. Based on the fur coat colour, the chimerism is judged to be (G) high if the mice are predominantly agouti and (H) mid to low if they are mostly black. (I) A typical F1 generation litter from mating between a high chimerism male and C57BL/6 female. Agouti pups are of ES cell origin while black pups originate from host blastocyst.

3.1.5. Producing the ROSA26-rTetR-KRAB(Kid1) mouse strain

A targeting strategy was devised to generate the rTetR-KRAB(Kid1) expressing mouse strain (3.1.5A). To achieve this, the rTetR-KRAB(Kid1) expression cassette was inserted into the ROSA26 gene locus. This locus was chosen as the integration site because it is a safe harbour locus. In the original mouse gene trap experiments, expression from this locus was ubiquitous¹⁸⁵ and did not cause any insertion side effects¹⁸⁶. My selected target site is close (280 bp upstream) to the provirus integration site described in these experiments. The replacement vector consisted of the rTetR-KRAB(Kid1) expression cassette flanked by 5'- and 3'-homology arms, which correspond to the region surrounding the target site in the wildtype locus. The 5' and 3' homology arms were 0.7 kb and 2.8 kb in length respectively. The exogenous CAG promoter was included in the expression cassette to drive rTetR-KRAB(Kid1) expression. This promoter was selected because it can achieve high expression levels¹⁸⁹, showing 8–10-fold higher expression than the endogenous ROSA26 promoter^{190, 191}.

Gene targeting was performed in ES cells as described earlier (Section 3.1.3). Clones with the desired modification were identified by PCR screening using two sets of primers (Figure 3.1.5A). The first set (SR1 and SR2) amplified a 1.1 kb stretch from outside the 5' homology arm to the expression cassette. The second set (SR3 and SR4) amplified a 2.2 kb stretch from the expression cassette to outside the 3' homology arm. Out of the 24 clones that were recovered, only two were positive for both PCR reactions (Figure 3.1.5B and Table 3.1.3).

Figure 3.1.5. Generation of ROSA26-rTetR-KRAB(Kid1) transgenic mouse strain.

(A) Targeting scheme to insert the rTetR-KRAB(Kid1) expression cassette into the ROSA26 gene locus (drawn approximately to scale). The sgRNA target site is in the first intron of the ROSA26 locus. The XbaI restriction site marks the location that is frequently targeted by other researchers. The replacement vector contains the rTetR-KRAB(Kid1) expression cassette flanked by 5'- and 3'-homology arms. An exogenous CAG promoter was included in the expression cassette to drive rTetR-KRAB(Kid1) expression. When CRISPR/Cas9-mediated double-stranded DNA break is repaired by homologous recombination, the rTetR-KRAB(Kid1) expression cassette is inserted into the Rosa26 locus. The positions of primers that will be used for screening ES cell clones and end-point reverse transcriptase PCR are indicated in the schematic. (B) PCR screening to identify targeted ES cell clones. Two separate PCR reactions are performed and the positions of the amplicons, with respect to the targeted ROSA26-rTetR-KRAB(Kid1) allele, are provided in A. Only results for positive clones are shown. The positive

control (+) was DNA from bulk GFP⁺ ES cells that were transfected with both the pSpCas9(Chimeric_BB)-2A-GFP and replacement vectors. Clone 16 produced germline-transmitting chimeras that were used to establish the ROSA26-rTetR-KRAB(Kid1) mouse strain. **(C)** Multiplex PCR-based genotyping strategy to distinguish between wildtype, heterozygous and homozygous mice. GR1 and GR2 are common primers that anneal to both the wildtype and knock-in alleles. However the knock-in allele is not amplified because the distance between the two primers is too far apart for it to be amplified during the short PCR extension time. GR3 primer is specific for the knock-in allele. **(D)** Genotyping results for wildtype, heterozygous and homozygous mice by multiplex PCR using primers GR1, GR2 and GR3. **(E)** End-point reverse transcriptase PCR to determine expression pattern of rTetR-KRAB(Kid1) mRNA in wildtype and heterozygous mice. The location of primers with reference to the targeted locus are shown in **A**. HPRT is used as an amplification and loading control. Abbreviations (in alphabetical order): ES cell embryonic stem cell; gDNA genomic DNA; sgRNA single guide RNA; HA homology arm; Het heterozygote; Hom homozygote; KI knock-in; pA poly(A) tail; RT reverse transcriptase; RT-PCR reverse transcriptase PCR; SI; small intestine; WT wildtype.

Chimeras were generated using ES cell clone 16, as described earlier (Section 3.1.4). Four chimeric founders were obtained (Appendix Figure A3.1.5). The three male chimeras had high ES cell contribution. The lone female chimera had low ES cell contribution, if any. All three male chimeras transmitted through the germline. Screening of F1 generation mice was performed using the same two sets of primers as the ES cell screen (Figure 3.1.5A).

To establish the ROSA26-rTetR-KRAB(Kid1) strain, heterozygous (ROSA26^{rTetR-KRAB/WT}) F1 generation mice from chimera 1 were backcrossed to C57BL/6 mice for at

least eight generations and subsequently maintained as a homozygous ($\text{ROSA26}^{\text{rTetR-KRAB/rTetR-KRAB}}$) line. Both heterozygous and homozygous mice were viable, fertile and did not display overt phenotypic or behavioural abnormalities. The transgenes were inherited, by succeeding generations, in Mendelian fashion.

For mice genotyping (of all subsequent generations after F1), I adopted a multiplex PCR-based method to differentiate between wildtype, heterozygous and homozygous mice (Figure 3.1.5C–D). The common primers GR1 and GR2 annealed to the 5'- and 3'-homology arms respectively on both the wildtype and knock-in alleles. Allele-specific primer GR3 annealed to rTetR-KRAB(Kid1) sequence on the knock-in allele. Amplification from primers GR1 and GR2 yielded a single 370 bp product that is specific for the wildtype allele. Although GR1 and GR2 can anneal to the knock-in allele, there was no amplification because the distance between the two primers (more than 3 kb) is out of range with the short thermocycling extension time used. Amplification from primers GR3 and GR2 yielded a 770 bp product that is specific for the knock-in allele.

For the rest of this thesis, I will refer to mice using the wildtype, heterozygous and homozygous designations. As an example: $\text{ROSA26-rTetR-KRAB(Kid1)}$ wildtype, heterozygous and homozygous indicate the $\text{ROSA26}^{\text{WT/WT}}$, $\text{ROSA26}^{\text{rTetR-KRAB/WT}}$ and $\text{ROSA26}^{\text{rTetR-KRAB/rTetR-KRAB}}$ genotypes respectively. Mice of the Tcf7l2-TRE strain will also be referenced in a similar manner.

Genes that are inserted into the ROSA26 locus normally display ubiquitous expression. To confirm that this is the case with the $\text{ROSA26-rTetR-KRAB(Kid1)}$ strain, end-point

reverse-transcriptase PCR (end-point RT-PCR) was carried out on various tissues. The primers (RR1 and RR2) annealed to the rTetR-KRAB(Kid1) sequence itself and amplified a 0.3 kb stretch. However the amplicon did not span an intron because there were none available. To ensure removal of contaminating DNA, RNA samples were treated twice with DNase I and reverse transcriptase negative controls reactions were included. For the small intestine and colon, the epithelial layer was processed separately from the rest of the mucosal and muscle layers. Heterozygous mice expressed rTetR-KRAB(Kid1) mRNA in a multiple tissues, including small intestine, colon, liver, spleen, kidney, lung, heart and brain (Figure 3.1.5E). Therefore rTetR-KRAB(Kid1) is broadly expressed in the ROSA26-rTetR-KRAB(Kid1) mouse.

3.1.6. Producing the Tcf7l2-TRE mouse strain

The mouse *Tcf7l2* gene has 17 exons, with numerous alternatively spliced exons (Figure 3.1.6A). A targeting strategy was designed generate the Tcf7l2-TRE mouse strain (Figure 3.1.6B). To achieve this, the TRE sequence was inserted into the *Tcf7l2* gene locus (3.1.6B). The target site was the third intron of *Tcf7l2*, but still within 3kb of the transcriptional start site. This is to ensure that when rTetR-KRAB(Kid1) binds to the TRE sequence, the KRAB repressor domain of rTetR-KRAB(Kid1) is well within the 15 kb working distance¹⁹². The replacement vector consisted of the TRE sequence flanked by 5'- and 3'-homology arms, which corresponded to the region surrounding the target site in the wildtype locus. The 5' and 3' homology arms were 0.7 kb and 1.0 kb in length respectively.

Figure 3.1.6. Generation of *Tcf7l2*-TRE transgenic mouse strain.

(A) The mouse *Tcf7l2* gene depicted as a schematic (drawn approximately to scale). The gene comprises of 17 exons (tall black boxes), with some of them being alternately spliced (tall white boxes). Key features are indicated: transcriptional start site (ATG) and stop codons. The relative positions of primers that will be used for semi-quantitative PCR are indicated in the schematic.

(B) Targeting scheme to insert the TRE sequence into *Tcf7l2* locus. The sgRNA target site is in the third intron of the *Tcf7l2* locus; but within 3 kb of the transcriptional start site. The replacement vector contains the 300 bp TRE sequence flanked by 5'- and 3'-homology arms. When CRISPR/Cas9-mediated double-stranded DNA break is repaired by homologous recombination, TRE is inserted into the *Tcf7l2* locus. The positions of primers that will be used

for screening ES cell clones are indicated in the schematic. **(C)** PCR screening to identify targeted ES cell clones. Three separate PCR reactions are performed and the positions of the amplicons, with respect to the targeted *Tcf7l1*-TRE allele, are provided in **B**. Only results for positive clones are shown. Clone 18 produced germline-transmitting chimeras that were used to establish the *Tcf7l2*-TRE mouse strain. **(D)** Multiplex PCR-based genotyping strategy to distinguish between wildtype, heterozygous and homozygous mice. The common primer GT2 anneals to the 3' homology arm. Allele-specific primers GT1 and GT3 anneal to the sgRNA target site (wildtype allele) and TRE sequence (knock-in allele) respectively. **(E)** Genotyping results for wildtype, heterozygous and homozygous mice by multiplex PCR using primers GT1, GT2 and GT3. **(F-G)** Semi-quantitative reverse transcriptase PCR to compare *Tcf7l2* expression and splicing between wildtype and homozygous mice. PCR was performed for 22, 25, 28, 31 or 34 cycles using cDNA as template. To check expression, primers RT1 and RT2 yield a single product. To check splicing, primers RT3 and RT4 yield two products because exon 4 is alternatively spliced. The relative positions of primers with reference to the gene locus are shown in **A**. HPRT is used as an amplification and loading control. Abbreviations (in alphabetical order): Col epi colon epithelial; ES cell embryonic stem cell; HA homology arm; Het heterozygote; Hom homozygote; KI knock-in; Pan pancreas; Semi-q RT-PCR semi-quantitative reverse transcriptase PCR; Sp spleen; Thy thymus; UTR untranslated region; WT wildtype.

Gene targeting was performed in ES cells as described earlier (Section 3.1.3). Clones with the desired modification were identified by PCR screening using three separate reactions (Figure 3.1.6B). The first reaction (ST1 and ST2) amplified a 1.1 kb stretch from outside the 5' homology arm to TRE. The second reaction (ST3 and ST4) amplified a 1.3 kb stretch from the TRE to outside the 3' homology arm. The third reaction (ST1 and ST4) amplified across both homology arms. This reaction yielded two possible amplicons: a 2.1 kb amplicon that was specific for the wildtype allele and a 2.4 kb amplicon that was specific for the knock-in. Out of the 24 clones that were

screened, six clones were positive for all three PCR reactions (Figure 3.1.6C and Table 3.1.3). It is interesting to note that clone 35 was negative for the 2.4 kb amplicon (primers ST1 and ST4) that is specific for the knock-in allele, despite having the amplicons for both the 5'-homology arm (ST1 and ST2) and 3'-homology arm (ST3 and ST4). A possibility is that the whole replacement vector, including the backbone, integrated in a concatemeric fashion¹⁹³. In such a scenario, the distance between the two flanking primers (ST1 and ST4) will be too far away to amplify (1 min thermocycling extension time) so the knock-in specific amplicon is absent.

Chimeras were generated using ES cell clone 18, as described earlier (Section 3.1.4). Four chimeric founders were obtained (Appendix Figure A3.1.6). All the chimeras were males and had high ES cell contribution. Only three of the chimeras transmitted through the germline. Chimera 4 was suspected to be a hermaphrodite because it was unable to sire pups. This was confirmed by dissecting the mice; the right-side reproductive organ was looked similar to an ovary and oviduct while the left-side reproductive organ was grotesquely enlarged and of undetermined gender. Screening of F1 generation mice was performed using the same three sets of primers as the ES cell screen (Figure 3.1.6B).

To establish the Tcf7l2-TRE strain, heterozygous F1 generation mice from chimera 1 were backcrossed to C57BL/6 mice for at least ten generations and subsequently maintained as a homozygous line. Heterozygous and homozygous mice were viable, fertile and did not display overt phenotypic or behavioural abnormalities. The transgenes were inherited, by succeeding generations, in Mendelian fashion.

For mice genotyping (of all subsequent generations after F1), I adopted a multiplex PCR-based method to differentiate between wildtype, heterozygous and homozygous

mice (Figure 3.1.6D–E). The common primer GT2 annealed to the 3'-homology arm on both the wildtype and knock-in alleles. Allele-specific primers GT1 and GT3 annealed to sgRNA target site (wildtype allele) and TRE sequence (knock-in allele) respectively. Amplification from primers GT1 and GT2 yielded a single 650 bp product that was specific for the wildtype allele. Amplification from primers GT3 and GT2 yielded a 950 bp product that was specific for the knock-in allele.

Inserting an exogenous sequence into the genome, even if the target site is an intron, could cause changes to gene expression and splicing. To verify that *Tcf7l2* expression and splicing remain unaffected, semi-quantitative RT-PCR was carried out using wildtype and homozygous *Tcf7l2*-TRE mice (Figure 3.1.6F–G). To ensure detection of all splice variants, the RT-PCR primers were located at the 5' end of the mRNA. The 3' end is unsuitable because there is a lot of alternative splicing in that region. To check expression, semi-quantitative RT-PCR was performed using primers that are specific for exon 1 (RT1) and exon 3 (RT2) (Figure 3.1.6A). There was no obvious difference in *Tcf7l2* mRNA expression between the homozygous and wildtype mice, in the spleen, thymus, pancreas and colon epithelial (Figure 3.1.6F). To check splicing, semi-quantitative RT-PCR was performed using primers that are specific for exon 3 (RT3) and exon 5 (RT4) (Figure 3.1.6A). This reaction yielded two products since exon 4 is alternatively spliced. Both isoforms were present in the spleen, thymus, pancreas and colon epithelial of homozygous and wildtype mice (Figure 3.1.6G). Moreover the relative proportion between the two isoforms remained unchanged between the two mice. I have shown that expression and splicing of the *Tcf7l2* gene is not affected in the *Tcf7l2*-TRE mouse.

3.2. Developing a Doxycycline Treatment Regimen

3.2.1. Induction kinetics of rTetR-KRAB(Kid1) *in vitro*

Now that both mouse strains were created, I could conduct experiments to better understand our Tet-inducible system and develop a treatment regimen. First, the induction kinetics of rTetR-KRAB(Kid1) was studied by using an *in vitro* reporter gene assay (Figure 3.2.1.A). For this assay, hematopoietic bone marrow cells were isolated from wildtype and heterozygous ROSA26-rTetR-KRAB(Kid1) mice and transduced with NUP98-HoxB4 fusion protein construct. This immortalised the bone marrow progenitor cells^{194, 195} so that they could be further manipulated and cultured *in vitro*. Subsequently a TRE-IRES-GFP reporter cassette, flanked by viral long terminal repeat (LTR) sequences, was stably integrated and GFP⁺ cells were isolated by FACS. The retroviral promoter in the 5'-LTR drove GFP expression, but expression levels could be modulated through the TRE sequence. The GFP⁺ lines had a purity of 94% to 97% (Appendix Figure A3.2.1).

The GFP reporter lines were cultured in medium containing varying concentrations of doxycycline and GFP signal was measured over a period of 10 days (Figure 3.2.1B–C). At the two highest concentrations of doxycycline (1 $\mu\text{g ml}^{-1}$ and 0.8 $\mu\text{g ml}^{-1}$), median GFP intensity was fully repressed by day 4. It took longer to achieve complete repression at lower doxycycline concentrations. In fact at 0.1 $\mu\text{g ml}^{-1}$ doxycycline, some GFP⁺ cells were still present at day 10. This suggests that partial repression can be achieved with careful doxycycline titration.

I have shown that the rTetR-KRAB(Kid1) repressor is functional. It can mediate gene repression when it is tethered in close proximity (via the TRE sequence) to the gene's promoter.

Figure 3.2.1. Induction kinetics of rTetR-KRAB(Kid1) *in vitro*.

(A) Schematic diagram of the experimental set-up. Induction kinetics was studied using immortalised GFP reporter bone marrow progenitors cells derived from wildtype and heterozygous ROSA26-rTetR-KRAB(Kid1) mouse. Following isolation from the bone marrow of adult mice, hematopoietic precursor cells were immortalised by retroviral transduction of a NUP98-HoxB4 fusion construct. A second transduction was performed to stably introduce the TRE-IRES-GFP reporter cassette. Pure populations of either GFP⁺ or GFP⁻ cells were obtained by fluorescence-activated cell sorting (FACS). Cells were treated with doxycycline and GFP signal was measured by flow cytometry. (B) Selected histogram plots tracking the GFP signal of doxycycline-treated GFP reporter cells (heterozygous line), at different time intervals. Data presented are raw values from a representative experiment. (C) Median GFP intensity of

doxycycline-treated GFP reporter cell lines over a period of ten days. The GFP intensities from the heterozygous line were normalised twice: first to the untreated control ($0 \mu\text{g ml}^{-1}$) and second to wildtype cells cultured in the same doxycycline concentration. Base line value was set using GFP⁻ cells. Data are plotted as mean $\pm$ SD from three independent experiments ($n = 1$ because each bone marrow line was derived from a single mice). Abbreviations (in alphabetical order): BM bone marrow; Dox doxycycline; Het heterozygote; IRES internal ribosome entry site; LTR long terminal repeat; WT wildtype.

3.2.2. Developing a doxycycline treatment regimen for mice

When administering doxycycline to induce rTetR-KRAB(Kid1) *in vivo*, pharmacokinetics factors come into play. It becomes necessary to test different administration routes and doses in order to develop a suitable regimen. To this end, I used bone marrow chimeras, which were reconstituted with the GFP reporter bone marrow cell lines created in the previous section (Section 3.2.1). The benefit of this approach is that the doxycycline response could be tracked periodically without having to sacrifice the experimental mouse, simply by taking blood samples. After lethal irradiation and hematopoietic reconstitution with the GFP reporter bone marrow line derived from heterozygous ROSA26-rTetR-KRAB(Kid1) strain (Figure 3.2.2A), mice were placed under different doxycycline treatments.

Common ways to administer doxycycline include non-invasive methods such as drinking water, food chow and gavage. Invasive intra-peritoneal (IP) injection is sometimes used. My initial plan was to deliver doxycycline by food chow (*ad libitum*) for sustained and long-term repression, while boosting initial doxycycline uptake by giving two or three daily shots of IP injections at the start of the treatment (Figure

3.2.2B). Two different doses of IP injection were administered: $25 \mu\text{g g}^{-1}$ body weight (low dose) and $50 \mu\text{g g}^{-1}$ body weight (high dose). For consistency, the base diet contained 1 g kg^{-1} doxycycline (high dose) in all treatments. Blood from treated mice were sampled at different time points to check the percentage of GFP⁺ cells (Figure 3.2.2C). When doxycycline was given only in the food, the percentage of GFP⁺ cells decreased to 40% by day 5 and complete GFP repression was achieved by day 10. This same trend was observed when mice were given IP injections in conjunction with doxycycline food. This showed that the additional IP injections did not boost GFP repression. In fact, mice that were given only daily low dose IP injections had a delayed response — only a slight decrease in the percentage of GFP⁺ cells was observed at day 10. This experiment was discontinued at day 10 because the mice, treated with daily IP injections, started to display signs of distress and apprehension during doxycycline administration.

Since IP injections did not have the booster effect that I had hoped for, I decided to continue with the food only administration route. Another two concentrations of doxycycline food (low dose 0.2 g kg^{-1} and mid dose 0.625 g kg^{-1}) were tested to find the lowest dose that would give fast repression. Gating was performed to determine the percentage of GFP⁺ cells in the major blood and spleen hematopoietic cell populations (Appendix Figure 3.2.2A–B). At all three doses, GFP⁺ cells were completely repressed in blood T cells and B cells by day 20 (Figure 3.2.2D). However for blood granulocytes and monocytes, only a slight repression was observed with the low dose food. This shows that it is possible to achieve partial repression *in vivo* by treating with food containing sub-optimal concentrations of doxycycline. In all the major spleen populations, better repression was achieved with doxycycline food containing increasing doxycycline concentrations (Figure 3.2.2E). This suggests that the doxycycline bioavailability in the spleen is lower than in blood.

Figure 3.2.2. Developing a doxycycline treatment regimen for mice.

(A) Schematic diagram showing how reporter bone marrow chimeras are produced for *in vivo* studies to optimise doxycycline delivery route and dosage. Following isolation from the bone

marrow of ROSA26-rTetR-KRAB(Kid1) heterozygous adult mice, hematopoietic precursor cells were immortalised by retroviral transduction of a NUP98-HoxB4 fusion construct. A second transduction was performed to stably introduce the TRE-IRES-GFP reporter cassette, flanked by viral LTR sequences. GFP expression is driven by the retroviral promoter. A pure population of GFP⁺ cells was obtained by fluorescence-activated cell sorting (FACS) and used to reconstitute the hematopoietic system of lethally irradiated recipient mice. With these bone marrow chimeras on-hand, different doxycycline treatments options can be trialled. **(B)** Experimental timeline with doxycycline treatment options. Two doxycycline delivery routes were tested: (1) food pellets and (2) intra-peritoneal (IP) injection. The idea is to deliver doxycycline in the food for sustained long-term repression, while boosting initial doxycycline uptake by giving several shots of IP injections at the start of the treatment. Treatment efficiency was determined by tracking the percentage of GFP⁺ cells in the blood over the course of the treatment. At the experimental end point (day 20), the mice were sacrificed and the spleen was also harvested. **(C)** Treatment of bone marrow chimeras with doxycycline food and booster IP injections. Data are presented as percentage change in GFP⁺ cells (normalised to day 0). The treatment regimens and number of mice used in each experimental group are described in the figure legend. Data plotted are mean $\pm$ SEM. The experiment was discontinued at day 10. **(D-E)** Treatment of bone marrow chimeras with food containing low, mid or high concentrations of doxycycline. **(D)** The percentage change in GFP⁺ cells in total blood peripheral blood mononuclear cells (PBMC) and the major blood subpopulations over the course of the experiment. Data plotted are mean $\pm$ SEM **(E)** The Percentage of GFP⁺ cells in total spleen cells and major splenic subpopulations at day 20. Data plotted are mean $\pm$ SEM. Abbreviations (in alphabetical order): BM bone marrow; Dox doxycycline; Het heterozygote; IP intra-peritoneal.

It is noted that a small percentage of blood granulocytes and monocytes, but not T cells and B cells, remained GFP⁺ despite persistent and continued treatment with doxycycline food (Figure 3.2.2D). Different levels of repression were seen among the spleen cell populations when comparing between the same dose (Figure 3.2.2E). A possible reason for this is variable rTetR-KRAB(Kid1) expression levels in different cell types. Takiguchi and colleagues¹⁹⁶ have shown that induction by tetracycline trans-activators in mice differ across cells of the hematopoietic system.

Based on these results, I have decided to go with the mid dose food for future *in vivo* experiments. My interest lies in the small intestine and colon because there is evidence that Tcf7l2 plays a role in intestinal homeostasis and cancer. I reckoned that the mid dose would be sufficient because doxycycline is absorbed by the gastro-intestinal track so these organs will have the highest bioavailability.

3.2.3. Repression of *Tcf7l2* in double transgenic mouse

To demonstrate that *Tcf7l2* can be repressed in mice with the Tet-inducible system, the ROSA26-rTetR-KRAB(Kid1) and Tcf7l2-TRE strains were inter-crossed to obtain double transgenic mice. To simplify comparisons between the groups of mice, all doxycycline-treated mice were heterozygous for the ROSA26-rTetR-KRAB(Kid1) allele and either wildtype, heterozygous or homozygous for the Tcf7l2-TRE allele (Figure 3.2.3B). Littermates of various other genotype configurations were used as untreated controls. Adult mice were fed mid dose doxycycline food (0.625 g kg⁻¹) *ad libitum*. After five days of treatment, the animals were sacrificed and the levels of *Tcf7l2* mRNA in the small intestine epithelial, colon epithelial and spleen were

quantified (Figure 3.2.3A). These organs were chosen because they express high levels of *Tcf7l2* mRNA and are relatively straightforward to harvest. Selective enrichment of the intestinal epithelium was performed because the Lgr5⁺ stem cells, which are reliant on Wnt signalling for proliferation, reside in this compartment.

In homozygous mice, *Tcf7l2* mRNA was completely repressed in the small intestine epithelial and colon epithelial by day 5 (Figure 3.2.3C–D). Heterozygous mice retained about 50% expression, which would be the case if only one allele is turned off. In the spleen, *Tcf7l2* levels were halved in homozygous mice compared to wildtype mice. (Figure 3.2.3E). This was expected because the earlier bone marrow chimeras experiment suggested that doxycycline bioavailability was lower in the spleen than in the blood.

I have shown that *Tcf7l2* can be repressed in double transgenic mice when they are given doxycycline food. This is a useful tool to study *Tcf7l2* gene function in various settings, especially in the intestine. This is because *Tcf7l2* is the main transcription factor that acts downstream of the canonical Wnt signalling pathway in the intestine¹⁵⁰. Although it has been implicated in the maintenance of the intestinal epithelium, its exact role is still controversial.

With this inducible *Tcf7l2* double transgenic mice on hand, the logical progression would be to study intestinal *Tcf7l2* function in vivo. However at this point in time, there are a limited number of double transgenic mice for experimental purposes because the mice colony is undergoing expansion. In the interim, I have derived organoid lines from the small intestine of double transgenic mice to study *Tcf7l2* function in vitro.

Figure 3.2.3. Repression of *Tcf7l2* in double transgenic mice.

(A) Schematic diagram showing the experimental timeline. Double transgenic mice were fed doxycycline food pellets (mid dose) *ad libitum* for a period of five days and their organs were harvested to check *Tcf7l2* repression. (B) Genotype configurations of double transgenic mice for each experimental group. Untreated mice having all possible configurations of the ROSA26-rTetR-KRAB(Kid1) and *Tcf7l2*-TRE alleles are used. All doxycycline treated mice are heterozygous for the ROSA26-rTetR-KRAB(Kid1) allele. The genotype at the *Tcf7l2* gene locus corresponds to the name of the experimental group. So the “WT”, “Het” and “Hom” groups are *Tcf7l2*-TRE wildtype, heterozygous and homozygous respectively. (C–E) Relative *Tcf7l2* mRNA expression, after five days of doxycycline treatment, in the (C) small intestine epithelial, (D) colon epithelial and (E) spleen. Quantitative reverse transcriptase PCR was performed using primers RT1 and RT2 that are specific for the first and third exons of the *Tcf7l2* gene respectively (please refer to Figure 3.1.6A for location of primers). Relative expressions were calculated using the comparative C_T method ($2^{-\Delta\Delta C_T}$). Data are plotted as mean

± SEM (Untreated $n = 7$; Wildtype $n = 7$; Heterozygous $n = 9$; Homozygous $n = 6$). Each dot represents individual mice. Statistical significance was determined with one-way ANOVA analysis follow by Tukey post-hoc test. ns: not significant, *: $p\text{-value} \leq 0.05$, ***: $p\text{-value} \leq 0.001$, ****: $p\text{-value} \leq 0.0001$. Abbreviations (in alphabetical order): Dox doxycycline; Het heterozygous; Hom homozygous; WT wildtype.

With intestinal organoids, it is possible to model the intestinal epithelium in vitro. When the $Lgr5^+$ stem cells in the intestinal epithelial crypt are cultured under appropriate conditions, they expand and self-organise in vitro into 3D-structures consisting of all epithelial layer cell subtypes. In order to maintain the stem cell niche, and ensure continued healthy growth of the organoids, the culture medium is supplemented with mesenchyme-secreted factors like EGF, R-spondin1 and noggin. The organoids highly resemble the physiological intestinal epithelium and has proven to be a useful tool to model adult stem cell differentiation⁴². Therefore the findings made in vitro would give an inkling of the processes in vivo.

3.3. Role of Tcf7l2 in the small intestine

3.3.1. Tcf7l2 in essential for the maintenance of small intestine organoids

The small intestine organoids were established from double transgenic mice, which were heterozygous for ROSA26-rTetR-KRAB(Kid1) and either wildtype, heterozygous or homozygous for Tcf7l2-TRE. Three days after passage, and after organoids had started to bud, doxycycline was added to the culture medium and the organoids were observed. (Figure 3.3.1A).

In homozygous organoids, complete *Tcf7l2* repression resulted in a stark loss of budding from the central lumen (Figure 3.3.1B–C). This is unlike wildtype organoids, which displayed healthy morphology with abundant budding. Budding was present but at reduced numbers in heterozygous organoids. Organoid growth was quantified by two-dimensional area (Figure 3.3.1D) and a similar trend was observed. On day three, heterozygous organoids were slightly smaller than wildtype organoids. This difference in area was significant ($p\text{-value} \leq 0.05$). Most strikingly, homozygous organoids had shrunk in area on doxycycline treatment. The absence of budding hints at the loss of the EphB/ephrinB gradient, which is responsible for positioning paneth cells at the crypt base¹⁴⁸ and promoting cell proliferation¹⁹⁷. Taken together, the results show that *Tcf7l2* is important for maintenance of small intestine organoids.

3.3.2. Repression of *Tcf7l2* rescues cystic APC^{min/+} small intestine organoids

To study the role of *Tcf7l2* in early tumour formation, organoids were established from the small intestine of adult APC^{min/+} mice, which were also homozygous for both ROSA26-rTetR-KRAB(Kid1) and *Tcf7l2*-TRE. The organoids initially displayed budding and were indistinguishable from healthy organoids, but spontaneously developed cystic morphology after several passages. This was congruent to the loss of heterozygosity of the APC gene^{198, 199}, correlating with the adenoma stage *in vivo*. I selected for cystic organoids by removing R-spondin-1 from the growth medium. This was possible because the Wnt signalling pathway was hyper-activated as a result of non-functional APC. The steps taken to achieve a pure culture of cystic organoids for subsequent experiments, are summarised in Figure 3.3.2A.

Figure 3.3.1. *Tcf7l2* is important for growth of small intestine organoids.

(A) The experimental plan for studying the significance of *Tcf7l2* in small intestine organoids. Organoids were established from the small intestine of double transgenic mice. All mice were heterozygous for the ROSA26-rTetR-KRAB(Kid1) allele and either wildtype, heterozygous and homozygous for the *Tcf7l2*-TRE allele. Once established, the organoids lines were cultured in medium with or without doxycycline (1 $\mu\text{g ml}^{-1}$) for a period of three days. Photographs were captured daily and RNA was harvested on day 0 and 2 to confirm that *Tcf7l2* is repressed. (B) Representative photographs of organoids over the course of the experiment. For each condition, the same organoid followed over time. (C) Quantification of organoid growth by number of

buds from the central lumen. On day three of doxycycline treatment, organoids were scored semi-quantitatively into four categories: no buds (0), a few buds (1 – 5), many buds (6 – 20) and too many to count (> 20). At least 30 organoids were scored for each treatment group and the counts are represented as a histogram (x-axis is the four categories so it is discontinuous). Data presented is from a representative experiment. **(D)** Quantification of organoid growth by 2-dimensional-area (2D-area). The 2D-areas of at least 30 organoids were measured for each group. The means for each group were normalised to the untreated control of the same genotype. Data are mean $\pm$ SD from two individual experiments ($n = 1$ because each organoid line was derived from a single mouse). Statistical significance was calculated using two-way ANOVA with Bonferroni post-hoc test. The significance values, annotated under the plots, are for comparisons between that treatment group and the wildtype group for that day. ns: not significant, *: $p\text{-value} \leq 0.05$, ****: $p\text{-value} \leq 0.0001$. Abbreviations (in alphabetical order): Dox doxycycline; WT wildtype; Het heterozygote; Hom homozygote.

On treatment with doxycycline, and hence *Tcf7l2* repression, some of cystic organoids reverted to budding morphology while the rest of the organoids had attenuated growth (Figure 3.3.2B–D). This is unlike untreated organoids, which stayed as cysts and grew in size. This shows that *Tcf7l2* is important for the persistence of adenomas and that it probably potentiates Wnt signalling in the small intestine.

Figure 3.3.2. *Tcf7l2* is important for adenoma formation in $APC^{min/-}$ organoids.

(A) Schematic diagram showing the derivation of cystic $APC^{min/-}$ organoids which correspond to the adenoma stage in intestinal cancer. Crypts from the small intestine of $APC^{min/+}$ mouse, which was homozygous for both the ROSA26-rTetR-KRAB(Kid1) and *Tcf7l2*-TRE alleles, were cultured in complete medium containing all essential growth factors. Initially, all crypts grew into organoids with the typical budding morphology. However spontaneous loss of heterozygosity of the APC gene resulted in fast-growing $APC^{min/-}$ organoids with cystic morphology. The cystic organoids were allowed to overtake the culture by passaging three times. On the third passage, budding organoids were eliminated by removing R-spondin1 from the culture medium. This resulted in a culture of purely cystic $APC^{min/-}$ organoids. (B) The experimental plan for studying the significance of *Tcf7l2* in cystic $APC^{min/-}$ small intestine organoids. The cystic organoids were cultured in medium with or without doxycycline ($1 \mu g ml^{-1}$

¹) for a period of five days. Photographs were captured daily. **(C)** Representative photographs of organoids over the course of the experiment. Red and blue arrows track organoids that reverted to budding morphology on doxycycline treatment. Photograph outlined in red shows one of the revertant at a higher magnification to emphasise the budding morphology. **(D)** Quantification of revertant organoids by number of buds from the central lumen. On day five of doxycycline treatment, organoids were scored semi-quantitatively into four categories based on the number of buds, and represented as a histogram (x-axis is the four categories so it is discontinuous). At least 30 organoids were scored for each treatment group. Data are mean $\pm$ SD from three independent experiments ($n = 1$ because the organoid line was derived from a single mouse). Abbreviation: Dox doxycycline.

CHAPTER 4

Discussion

4.1. A Platform for Inducible Gene Repression in Mice

4.1.1. A choice between Tet-inducible activation or repression

The Tet-inducible system is a powerful tool for regulating gene expression. In most applications of this system, transcriptional activation of a target gene is accomplished by tethering a trans-activator close to the gene's transcriptional start site via the TRE sequence. This bi-partite trans-activator is a fusion between the VP16 transcriptional activation domain and either TetR (binds TRE in the absence of doxycycline) or rTetR (binds TRE in the presence of doxycycline). Deuschle and colleagues⁹¹ were the first to develop a trans-silencer, by replacing the VP16 domain with the Krüppel-associated box (KRAB) transcriptional repressor domain of the Kox1 protein. The KRAB domain mediates transcriptional silencing over long distances of up to 15 kb^{192, 200}, by inducing heterochromatin spreading through recruitment of heterochromatin protein 1 β (HP1 β)¹⁹². The ability to confer repression is not limited to the KRAB domain. Ssn6 and Tup1 have separately been fused to TetR for transcriptional repression in *Saccharomyces cerevisiae*²⁰¹.

Trans-silencers were originally developed with the intention to use them, together with trans-activators, in order to reduce background activity in Tet-inducible activation

systems^{79, 80, 91}. However they are not relegated to that role. Trans-silencers can also be exploited to mediate repression of target genes, in a Tet-inducible manner, when they are tethered close to the gene's promoter^{92, 93, 201}.

Although both the activation and repression versions of the Tet-inducible system allow the user to regulate gene expression, the usage and outcomes are vastly different. With the repression system, the target gene is expressed at physiological level before doxycycline treatment. This is because the target gene remains under the control of the endogenous promoter. However, the gene can be repressed when desired. With the activation system, an exogenous cassette with the transgene is randomly inserted into the genome. Since the transgene is under the control of the TRE-minimal promoter, there is little or no expression if the trans-activator is not bound. But when the trans-activator binds, the transgene is overexpressed. Expression can be way beyond physiological levels because the exogenous viral VP16 domain is such a potent trans-activator, that it can promote expression levels 100-fold greater than other activation domains such as the human Sp1 and viral Tat1⁷⁷. Moreover, a truncated version of the VP16 activation domain (amino acids 410–490) amplifies the activity of the pancreas specific Pdx1 transcription factor by more than 8-fold in the liver, when the two proteins are fused together⁷⁸.

Given that overexpression or ectopic expression could lead to artefacts and false observations, the activation system is clearly at a disadvantage to the repression system for gene function studies. Therefore it comes as a surprise that the repression system is less widely utilised than the activation system. Perhaps the sparse use of the repression system in mice could be explained by the lack of available strains expressing the trans-silencer. A search on both the Jackson Laboratory (JAX) and European Mutant Mouse

Archive (EMMA) databases turned up only one hit for a trans-silencer strain. In contrast, there were numerous hits for trans-activator strains. Researchers might be opting for a ready-made trans-activator strain, rather than investing huge amounts of time and money to create new trans-silencer strains.

The repression system is excellent for studying gene function because it relies on knockdown rather than overexpression. Hence the platform, established in this work, is based on the Tet-inducible repression system.

4.1.2. Optimising the Tet-inducible trans-silencer for in vivo gene repression

Since Deuschle and colleagues⁹¹ were able to achieve successful transcriptional repression with their TetR-NLS-KRAB trans-silencer, I mimicked their design. The intervening NLS sequence was kept to ensure nuclear localisation, but two improvements were made.

The first improvement was to adapt the trans-silencer for in vivo use, by replacing the TetR domain with rTetR. The TetR-KRAB trans-silencer is not ideal because mice would have to be given doxycycline continuously, to maintain endogenous gene expression. As a consequence of prolonged administration, doxycycline would accumulate in bones^{202, 203}, by chelating calcium ions^{204, 205}, and in fats because of its high lipophilicity^{206, 207}. Even after doxycycline treatment has ceased, residual amounts of doxycycline would continue to leach from the bones and fats so repression would take a long time to kick in. This can be solved by using a rTetR-KRAB version of the

trans-silencer, since it binds TRE¹⁸⁰ and represses the gene only in the presence of doxycycline. Mice are treated with doxycycline when repression is desired.

The second improvement was to replace the full length Kox1 KRAB domain with the minimal Kid1 KRAB repression domain. The KRAB domains of Kox1 and Kid1 can be used interchangeably because both proteins are from the same zinc-finger protein family²⁰⁸ and they share high amino acid sequence conservation^{181, 208}. Kid1 was selected because the minimal KRAB domain, required for repression, is well defined¹⁸¹. The replacement shortens the length of the KRAB component from 121 to 42 amino acids and could help to avoid possible squelching effect. Typically, squelching occurs when excess transcriptional activator domain compete for limiting co-activators of the basal transcriptional machinery, leading to general and non-specific inhibition of transcription⁸². Theoretically, the reverse can also happen. Full length KRAB domain could possibly sequester limiting co-repressors and non-specifically affect repression of other genes. Removing extraneous sequences minimises the likelihood of this.

I have demonstrated that the improved rTetR-KRAB(Kid1) trans-silencer is inducible and functional by using *in vitro* and *in vivo* assays. With both bone marrow cell lines and chimeras, GFP was repressed on treatment with doxycycline. In theory, repression should be reversible when doxycycline treatment is withdrawn because rTetR-KRAB(Kid1) would be released from the TRE sequence. Although not investigated in this work, others have shown that KRAB-mediated repression is reversible after prolonged silencing, if the repression happened during the adult stage^{92, 209, 210}. It is irreversible if exerted during early embryonic development²¹⁰, in particular between embryonic days 3.5 to 7.5²⁰⁹. It is interesting to note that this window period coincides with the restoration of DNA methylation during embryonic epigenetic remodelling²¹¹.

Indeed, a trans-silencer with the KRAB domain induces stable DNA methylation in the embryonic stage²⁰⁹. This is distinct from the dynamic H3K9me3 heterochromatin modifications induced by the same trans-silencer in the adult stage²⁰⁹.

4.1.3. Potential Applications of this Platform

Animal studies remains one of the best methods for drug target validation because the complexity and interactions within the whole organism cannot be recapitulated *in vitro* or *ex vivo*. This platform is especially suited for this purpose because it not only allows inducible and reversible gene knockdown in mice, but it also mimics aspects of taking an oral drug; repression is dose-dependent and systemic. Partial repression of the target gene can be achieved either by using the heterozygous TRE strain or titrating the doxycycline dose. Since repression is systemic, it could provide information on how other processes or organs are affected by targeting that specific protein. The platform can also be used broadly, to study the function of the target gene in healthy and disease settings.

The ROSA26-rTetR-KRAB(Kid1) strain can be paired with TRE strains for repression of other target genes. Ideally for the repression to work, insertion of the TRE sequence must not interfere with expression or splicing of the gene. There is some flexibility in the location for inserting the TRE sequence since the KRAB repressor domain achieves silencing through epigenetic modifications over long distances, and not by steric hindrance. The process of creating new transgenic mouse strains has been greatly facilitated by advances in gene editing technology. With CRISPR/Cas9 technology, it is both feasible and manageable to create a new strain for each target gene.

4.2. Considerations for Creating Transgenic Mouse

4.2.1. A pipeline for creating transgenic mouse strains

This thesis describes a workable pipeline for creating transgenic mouse strains by using the CRISPR/Cas9 technology. Optimised parameters include a screening assay to determine cleavage efficiency conferred by sgRNAs and a strategy for CRISPR/Cas9-assisted targeting in ES cells. This pipeline has been utilised by my colleagues to create other transgenic strains. It could also serve as a guide for others who are interested in transgenesis work. In the subsequent sections, I will elaborate on the rationale behind specific choices and highlight some important aspects to consider when performing gene targeting with CRISPR/Cas9 technology.

4.2.2. Selecting an assay for screening potential sgRNAs

It is evident that an sgRNA's ability to direct Cas9 nuclease activity at the target site, is influenced by chromatin accessibility. There is a positive correlation between sgRNA activity and mRNA expression levels²¹². Furthermore Cas9-sgRNA complexes are unable to induce cleavage in heterochromatin DNA, but can do so when that same stretch of DNA is in euchromatin state⁹³. When the structural models of the Cas9-sgRNA complex and heterochromatin DNA were superimposed, it was noted that nucleosomes sterically hinder the interaction between the two components⁹³. Indeed, cleavage by the Cas9-sgRNA complex was blocked by artificially induced nucleosome assembly²¹³. However this could be rescued by removing those nucleosomes through the use of a truncated Snf2 chromatin remodelling enzyme²¹³.

Chromatin accessibility can be deduced based on information from RNA-seq databases, and web-based prediction tools, such as “CHOPCHOP”^{178, 179} and “CRISPR design”²¹⁴, can help with sgRNA selection. But ultimately, potential sgRNAs have to be screened. *In vitro* sgRNA screening methods are convenient and quick, but should be avoided because they do not provide accurate information on sgRNA activity²¹². When using such methods, Cas9 protein, in vitro transcribed sgRNA and DNA cleavage template are combined and the cleavage reaction is allowed to happen. Since the cleavage template is obtained by PCR amplification it lacks higher order chromatin structure. To account for chromatin accessibility, sgRNAs have to be screened in cells, preferably using the cells that are intended for the final gene targeting experiment.

4.2.3. Quantifying sgRNA efficiency

The Sanger sequencing assay for determining sgRNA activity, as described in this work, is quasi-quantitate. For exact quantification of sgRNA activity, an additional analysis can be performed using the web-based program TIDE (Tracking of Indels by Decomposition)²¹⁵. This algorithm measures the decay in a sequence chromatogram downstream of the cleavage site²¹⁵. Similarly, results from Surveyor and T7 endonuclease assays can be quantified by measuring the relative intensities of gel bands with ImageJ software or other similar programs.

The purpose of quantification is to differentiate between sgRNAs associated with high or low cleavage activity. This was not necessary for ES cell targeting, as good targeting efficiencies were attained even with randomly selected sgRNAs. Somatic cell gene

targeting is bound to be more difficult because the chromatin is less accessible than in ES cells²¹⁶⁻²¹⁸. Furthermore, somatic cells tend to repair double-stranded breaks by non-homologous end joining (NHEJ) rather than homologous recombination, which is opposite to ES cells^{219, 220}. For instance, the frequency of homology-directed repair in mouse embryonic fibroblast cells, ear fibroblast cells and mammary epithelial cells is 3-fold lower than ES cells, despite having similar double-stranded break efficiencies²¹⁹. Therefore it might be prudent to optimise every step along the process, including selecting an sgRNA with high cleavage activity, to boost the final targeting efficiency.

4.2.4. Minimising off-target effects

Unbiased genome-wide screens performed using deep sequencing, have shown that CRISPR/Cas9 gene editing can cause off-target mutations, albeit at low frequencies²²¹⁻²²⁵. This would be a safety concern if the technique were applied to gene therapy. If used for creating stable cell lines, it might cast doubt on whether an observed phenotype is due to an off-target mutation in an important genomic locus. For mouse transgenesis, this is less of a concern because mutations that may arise from off-target cleavage can be segregated by back-crossing, as was done for both of the strains created in this work.

There are several strategies to reduce off-target cleavage with CRISPR/Cas9 gene editing. An obvious method would be to select unique sgRNAs that do not have extensive sequence homology with other loci, except the intended site. To borrow a page from ZFN- and TALEN-assisted gene editing, paired Cas9 nickase can be used to create two single-stranded break on each DNA strand, at adjacent sites^{117, 226, 227}. Since two sgRNAs are required, one to guide each Cas9, it is unlikely that both sgRNA

sequences would occur close together at off-target loci. A variant of this strategy is to use two catalytically inactive Cas9 (dCas9) fused to FokI endonuclease^{223, 228}. Another strategy is to limit the time that Cas9 is present in the cells, by delivering purified Cas9 protein itself instead of Cas9 expression vector^{229, 230}. If Cas9 is present for a long time, even after it has fulfilled its purpose of on-target DNA cleavage, there is increased opportunity for it to cause off-target cleavage. Understanding the structure and interactions between Cas9, sgRNA and its target DNA can also provide clues on how to minimise off-target activity. Truncated sgRNAs of 17- to 18-nucleotide in length can reduce off-target activity. Possibly this works by reducing the free energy at the sgRNA-DNA interface so that mismatches between the sgRNA and DNA are not tolerated²³¹⁻²³³. These different strategies can be used in combination to further reduce the chances of off-target mutations.

4.2.5. Skewing cellular repair pathway for preferential DNA insertion

Inserting a DNA fragment into a specific locus with nuclease-assisted gene targeting involves two steps. First, a double-stranded break is created at the target site and this is followed by homology directed repair (HDR) using the supplied DNA template. It is important to make the distinction that the nuclease, be it ZFN, TALEN or Cas9, facilitates the double-stranded break and not homologous recombination repair. The outcome of gene targeting will still largely depend the prevalence of the non-homologous end joining (NHEJ) and HDR pathways in the cells. This is the limiting factor when trying for targeted insertion in somatic cells, because the HDR pathway is much less active than NHEJ pathway^{219, 220}. For instance when I attempted to target hybridoma cells using CRISPR/Cas9 technology, cleavage efficiency was high but

recombination frequency was extremely low (Unpublished data). With a similar CRISPR/Cas9 targeting scheme, Pongson and colleagues²³⁴ also observed low insertion efficiency (less than 1%) when targeting hybridoma cells, despite good cleavage efficiency.

It is possible to force the outcome to obtain targeted insertion, by blocking the NHEJ repair pathway. This can be done by treating with drugs that inhibit essential enzymes in the NHEJ repair pathway, such as Ku70 and DNA ligase IV^{235, 236}. It is unlikely that this increases the frequency of HDR per se. Instead it probably causes cells, which do not undergo HDR, to enter cell arrest and eventually apoptosis. Thereby enriching for cells that undergo HDR and have the targeted insertion.

4.3. Considerations for Developing a Doxycycline Regimen

4.3.1. Pros and cons of the reporter bone marrow chimera system

There are several advantages of using reporter bone marrow chimeras in the development of a doxycycline treatment regimen. The protocol for immortalising hematopoietic bone marrow progenitor cells^{194, 195} and creating bone marrow chimeras²³⁷⁻²³⁹ are well established. Once immortalised, the bone marrow cells can be maintained for some time *in vitro* without losing their stem cell potential and ability to differentiate down the different hematopoietic lineages when transplanted back into mice^{194, 240}. This window period, in which the cells can be cultured *in vitro*, is an opportunity to manipulate them. In this work, it was exploited to introduce the GFP reporter construct.

However this system has its limitations. First, only the hematopoietic system is reconstituted with the reporter cells. Therefore the repression kinetics in other organ systems has to be inferred. The fastest repression will probably occur in the gastrointestinal tract, kidneys and liver because these organs will have the highest doxycycline bioavailability. Like most tetracyclines, orally administered doxycycline is mainly absorbed by the duodenum^{241, 242}, metabolised in the liver^{206, 243} and excreted by the kidneys^{206, 243, 244}. There might also be repression in the brain since doxycycline is lipophilic and can cross the blood brain barrier²⁰⁶. However even at high doxycycline doses, repression in the brain would be nominal as doxycycline concentrations in the brain is 4 to 7-fold lower than serum^{245, 246}.

Second, repression is delayed because GFP is very stable, having a long half-life of 26 hours^{247, 248}. Most cellular proteins are less stable. Take for example in yeast, the modal protein half-life is 5.5 hours and more than 90% of proteins have a half-life of less than 8 hours²⁴⁹. Based on this, it is expected that most target genes will be repressed in a shorter time compared to the GFP reporter. It is possible to get a better sense of the repression kinetics by using a reporter gene with a half-life that matches cellular proteins. Viable alternative reporters include luciferase that has a half-life of 3 hours²⁵⁰ and destabilised versions of GFP that have half-lives ranging from 30 minutes to 2 hours^{251, 252}.

Third, it is not suitable for examining the reversibility of the Tet-inducible repression system. The hybrid retroviral LTR²⁵³ used to drive GFP expression in the bone marrow chimeras is active in many hematopoietic cell subsets. However exogenous retroviral promoters are prone to epigenetic transcriptional silencing in host cells^{189, 254}. The exact

mechanism is not well defined but may be similar to endogenous retroviral silencing during embryo development. This involves initial repressive histone H3K9me3 modification by KRAB-associated zinc finger protein and Kap1²⁵⁵, followed by stable *de novo* DNA methylation^{189, 210, 256}. It is possible that repression by the rTetR-KRAB trans-silencer recreates the initial histone modification and pre-disposes the locus to DNA methylation, thereby rendering the repression irreversible.

Despite the limitations of this system, it served as a convenient method for tracking repression over the course of doxycycline treatment. This allowed me to test different administration routes and doses to find a suitable treatment regimen.

4.3.2. Balancing multiple factors to find a suitable doxycycline regimen

Any of the tetracyclines would be able to induce the Tet-inducible systems. However doxycycline is normally the drug of choice for Tet-inducible systems because it can readily penetrate multiple organs due to its lipophilic nature^{206, 207}, has a long serum half-life of 14 to 22 hours^{207, 244, 257} and is inexpensive. Therefore doxycycline was used in this work.

Considering that doxycycline has a long half-life, the interval between doses can be drawn-out. In fact, the recommended prescription for humans is orally at an initial dose of 200 mg on the first day, followed by daily doses of 100 mg²⁵⁸. This allows some flexibility in the dosing schedule and opens up multiple options for administering doxycycline in mice. The common routes in mice can be classified into invasive and non-invasive methods.

Invasive methods, such as orogastric gavage and intra-peritoneal injection, ensure consistent dose and it is feasible to administer daily. However this may not be ideal for long-term treatment because repeated administration may cause stress to the mice. Even a single round of gavage can cause elevated heart rate, blood pressure and corticosterone stress hormone²⁵⁹. The dosage with non-invasive methods, such as drinking water and food may be inconsistent because it depends on the mice's eating or drinking habits. However such methods are well tolerated and easy to administer. Out of the two non-invasive methods, mice are less receptive to doxycycline in the drinking water because it is unpalatable even when sucrose is added²⁶⁰. Taking all these factors into account, food was the logical choice for doxycycline administration. The final dosage selected is well within safety limits because it is less than 7% of the oral LD50 in mice²⁵⁸.

4.3.3. Doxycycline sensitivity and dosage

The rTetR version of the trans-silencer might be more suitable than the TetR version for *in vivo* use, but there is a trade off. The mutations to convert TetR to rTetR also diminish the doxycycline sensitivity by about 100-fold²⁶¹. This means that a high doxycycline dose has to be administered and it would also take longer to reach high enough levels in serum and tissue for trans-silencer activation. In retrospect, one of the improved versions of rTetR could have been used. These improved versions carry several amino acid mutations that restore sensitivity to doxycycline. For example, the improved rTetR-V10 and rTetR-V16 versions only require 44 ng ml⁻¹ and 9 ng ml⁻¹ doxycycline to reach maximum activity, respectively^{73, 262}. Not only are these

concentrations significantly lower than the 1000 ng ml⁻¹ required for comparable activity of rTetR-M2^{73, 262}, but they are also in the range required for the original TetR activation²⁶¹.

4.4. Understanding the role of Tcf7l2 in the intestine

4.4.1. Relevance of the inducible and reversible Tcf7l2 knockdown mouse

Although Tcf7l2 has been implicated in many homeostatic processes and diseases in adulthood, its exact roles are not well understood. This is partly due to a lack of suitable Tcf7l2 transgenic mouse strains. The transgenic mouse described here is the first instance of an inducible and reversible *Tcf7l2* knockdown mouse. Repression of Tcf7l2 can be controlled by administration or withdrawal of doxycycline. Since repression occurs ubiquitously throughout many tissues, the use of this transgenic mouse is not restricted to studying function or disease in one specific organ. Therefore it is a useful tool to shed light on the function of Tcf7l2 in many processes, especially multi-organ diseases such as colorectal cancer and type 2 diabetes. It is also entirely possible that Tcf7l2 has other novel functions yet to be uncovered. Not only is Tcf7l2 expressed across many tissues in humans²⁶³, but it also targets many genes involved in metabolic and cardiovascular disorders²⁶⁴.

An increased risk of type 2 diabetes is associated with a SNP (rs7903146) within intron 3 of *Tcf7l2*, as identified by genome-wide association studies (GWAS)^{265, 266}. There are differing views on how this relates to disease pathology. On one hand, pancreatic β -cells in subjects who carry the “T” risk allele, are less responsive to incretin hormones

than those who carry the “C” protective allele²⁶⁷. Incretin hormones help to maintain blood glucose levels after food intake, by augmenting β -cell insulin secretion. If β -cells become less responsive, chronic hyperglycemia could trigger β -cell apoptosis and eventually lead to type 2 diabetes^{268, 269}. On the other hand, the risk allele confers higher expression of Tcf7l2 in pancreatic islets compared to the protective allele^{270, 271}. This is counterintuitive because siRNA knockdown of Tcf7l2 reduces insulin secretion from β -cells^{271, 272}. This discrepancy suggests that the SNP may affect Tcf7l2 in ways other than just the expression level, for example alternative splicing. Numerous transcript variants arising from alternative splicing of Tcf7l2, especially in the 3' region, have been documented^{150, 273, 274}. However, few groups have looked into a possible link between the SNP and splicing. Osmark and colleagues²⁷⁴ investigated whether Tcf7l2 exon 4 usage was affected by the rs7903146 SNP in pancreatic islets. They did not observe appreciable difference in the abundance of exon 4-containing transcripts when they compared between donors with the “C/C” and “C/T” SNP genotypes²⁷⁴. The possibility that the rs7903146 SNP may affect splicing in the pancreas should not be ruled out, based on this experiment alone. This is because none of the donors were homozygous for the risk allele (“T/T” genotype) and this is only one of the several alternative exons used in the pancreas^{274, 275}. Another possible explanation for the discrepancy is that Tcf7l2 has a multifaceted role in the aetiology of type 2 diabetes, since it targets many genes that are involved in metabolic processes^{264, 271}.

Evidence for the involvement of Tcf7l2 in adult intestinal homeostasis and colorectal cancer is compelling. However there are still some inconsistencies to iron out. Since the Tcf7l2 mouse colony is still being established, I performed preliminary experiments using organoids derived from the small intestine of the Tcf7l2 mice. Intestinal organoids can provide useful insights because they closely mimic the epithelial micro-

environment. However it is still an artificial setting and there is no information about how the mesenchyme contributes to the general physiology. Therefore it is important to verify any findings *in vivo*. In the next two sections (Section 4.4.2 and 4.4.3), I will discuss the biological significance of the results from the organoid experiments and suggest how the inducible and reversible Tcf7l2 knockdown mouse can be used to address the issues and inconsistencies in the field.

4.4.2. The role of Tcf7l2 in intestinal homeostasis

There is no doubt that the Wnt signalling pathway is essential for maintenance and proliferation of the Lgr5⁺ stem cells in the gut^{143, 144, 151}. However there are conflicting views as to whether the downstream Tcf7l2 transcription factor behaves as an activator that propagates Wnt signalling, or as a repressor that regulates the level of Wnt signalling.

Using organoid cultures, I confirmed that Tcf7l2 is important for maintenance of small intestine epithelial cells. Complete Tcf7l2 repression resulted in the absence of budding and decrease in organoid size. This phenotype is similar to that in mice with attenuated Wnt signalling pathway^{143, 144, 151}. This pathway is the master regulator of cell proliferation and exerts regulation through multiple mechanisms. One way that it does this is by directly promoting c-Myc expression, which in turns suppresses the p21 cell cycle inhibitor²⁷⁶. The Wnt pathway also directly regulates the expression of EphB receptors, which are important for the EphB/ephrinB gradient along the intestinal crypt axis¹⁴⁸. Both EphB receptors and ephrinB ligands are membrane-bound, but they are expressed in an inverse pattern in the crypts. Cells at the base of crypts express EphB

receptors while those at the top express ephrinB ligands. When the receptor and ligand contact each other, a repulsive signal triggers cytoskeleton remodelling and causes the cells to move apart¹⁴⁹. This gradient is important for stem cell proliferation¹⁹⁷ and organisation of the stem cell niche in the gut¹⁴⁸. In EphB3 receptor knockout mouse, terminally differentiated paneth cells are scattered along the length of the crypts instead of confined at the crypt base¹⁴⁸. The loss of organisation would explain the lack of budding, and hence the slower growth in the small intestine organoids.

To substantiate these claims, follow-up experiments in organoids can be performed. The EphB/ephrinB gradient, or lack of it, can be visualised by immunohistochemistry with anti-EphB2 or anti-EphB3 antibodies for the receptor, and anti-ephrinB1 antibodies for the ligand. Tcf7l2 repression also affected proliferation and probably causes the stem cells to differentiate. Proliferation can be verified by Ki67 staining or BrdU labelling. The presence of stem cells can be detected with anti-Ascl2 antibody, which shows similar expression pattern to Lgr5²⁷⁷. It is not possible to identify stem cells with the *bona fide* Lgr5 stem cell marker in immunohistochemistry, because Lgr5 has a very low expression level. Signal strength is weak even when the anti-Lgr5 antibody is conjugated to Qdot probes, which are known to be exceptionally bright²⁷⁸.

To verify these findings *in vivo*, the inducible and reversible Tcf7l2 mouse can be treated with doxycycline for several days before examining the small intestine. Again, immunohistochemistry can be performed to examine if stem cells or the EphB/ephrinB gradient are lost and whether there is proliferation. Another possible consequence to look out for would be the integrity of the epithelial lining, since this is disrupted in Tcf7l2 knockout mice¹⁵¹. Leakiness of the epithelial lining can be determined using

functional assays, including measuring faecal albumin levels or the presence of an orally administered indigestible compound, like fluorescence-labelled dextrans, in the blood²⁷⁹.

The role of *Tcf7l2* in the colon remains ambiguous. Angus-Hill and colleagues¹⁷³ found that knockout of *Tcf7l2* in some colonic crypts caused hyper-proliferation in those crypts. However van Es and colleagues¹⁵³ observed an opposite effect when using a tamoxifen-inducible *Tcf7l2* knockout mouse; There was a lack of Ki67⁺ proliferating cells in the colonic crypts. This discrepancy can be attributed the differences in the two mouse models. In the model used by van Es and colleagues¹⁵³, exon 11 that corresponds to the HMG DNA binding domain, is deleted on tamoxifen induction. With this strategy there is still expression of a short isoform of *Tcf7l2* that terminates using the stop codon in exon 8²⁷³. This isoform contains the β -catenin binding domain but lacks the HMG DNA binding domain^{280, 281}. Like its longer counterparts, the short isoform is also able to propagate Wnt signalling but it activates a different subset of genes. It does so by binding to β -catenin and forming complexes with other transcription factors, such as SF-1 and C/EBP-alpha²⁸¹. In the model used by Angus-Hill and colleagues¹⁷³, exon 1 is deleted as early as the germline stage, but only in less than 1% of the crypts. It is uncertain if there are compensatory mechanisms due to germline deletion. The *Tcf7l2* mouse strain developed in this thesis has none of the complications arising from the models used by Angus-Hill and van Es^{153, 173}. Therefore it is ideal for clarify the role of *Tcf7l2* in the colon.

4.4.3. The role of Tcf7l2 in intestinal cancer

Using small intestine organoids derived from Apc^{Min} mouse, I have shown that the cystic morphology of Apc^{Min/-} organoids could be reverted to healthy-like budding by repressing Tcf7l2. Although there were a large number of organoids that remained cystic on Tcf7l2 repression, their proliferation was noticeably attenuated. There are two possible reasons for this. Either the repression period was not long enough for all the organoids to revert, or repression was so strong that the level of Wnt signalling was too low to sustain survival. If the reason were the later, the revertants could have a form of APC mutation that confer growth advantage, while the non-revertants with a different type of APC mutation lacked the growth advantage and would eventually undergo apoptosis. In any case, the results suggest that adenomatous small intestine epithelial cells are reliant on Tcf7l2 expression to maintain their transformed state. Obviously this is only in early stage adenomas and it remains to be seen whether Tcf7l2 repression will be therapeutic in the later stages of small intestine cancer. But the potential is there since the Wnt signalling pathway mediates many processes throughout cancer progression, from adenoma through carcinoma to metastasis²⁸²⁻²⁸⁵. In fact, restoration of APC expression and hence inhibition of Wnt signalling pathway, caused regression in advance colon carcinomas in mice¹⁶⁵. These carcinomas were artificially created by knocking down Apc and mutating both Kras and p53¹⁶⁵, which are three of the four key driver genes in colon cancer^{286, 287}.

The role of Tcf7l2 in colorectal is not so clear-cut. Due to the possibility of Tcf7l2 being important for stem cell proliferation in healthy colon, it would make sense if Tcf7l2 acts as an oncogene to promote colorectal cancer. However multiple studies provide evidence in favour of a tumour suppressor role. A profile of the genetic changes

that occur during colon cancer found sporadic cases with inactivating Tcf7l2 mutations²⁸⁷. In addition, Tcf7l2 is expressed at lower levels in colon cancer tissues compared to healthy colon tissues from the same patient¹⁵⁰. If Tcf7l2 were an oncogene instead of a tumour suppressor, it should be functional and upregulated in colon cancer biopsies. There is also direct evidence that Tcf7l2 acts as a tumour suppressor. In *Apc*^{min/+} mice, Tcf7l2 haploinsufficiency resulted in a higher number of tumours that showed characteristics of the late adenoma stage¹⁷³. Moreover, Tcf7l2 knockdown enhanced growth of colorectal cells¹⁷².

To study the relevance of Tcf7l2 in colorectal cancer and at different stages throughout cancer progression *in vivo*, the *Apc*^{Min} strain can be mated to the inducible Tcf7l2 strain. *Apc*^{Min/+} mice develop spontaneous adenomas mostly in small intestine and less frequently in the colon^{147, 288}. However these adenomas rarely become invasive, and mice die from an overwhelming number of adenomas that cause anaemia and obstruction of the small intestine^{147, 288}. Therefore to model later stage cancer, it is necessary to introduce environmental stressors such as dextran sodium sulphate or bile acid deoxycholate^{289, 290}. Once late stage tumours develop, Tcf7l2 can be repressed before determining the number and stage of tumours. Tumour staging can be performed by histological methods such as hematoxylin and eosin stain.

4.4.4. Targeting the Wnt signalling pathway through Tcf7l2 as a possible cancer therapy

Elevated Wnt signalling correlates with poor prognosis in many types of cancers. Therefore targeting the Wnt signalling pathway might be a promising treatment option

for such cancers. There are several drugs in clinical trials that act directly on the Wnt signalling pathway. Many of these drugs target upstream components, for example Porcupine inhibitors (LGK974 and ETC-159)^{166, 167}, anti-Frizzled receptor antibody (OMP-18R5)¹⁶⁸ and anti-R-spondin 3 antibody (OMP-131R10)²⁹¹. But there are some drugs that target downstream components, for example by preventing β -catenin from interacting with transcriptional co-activators (PRI-724)²⁹². Regardless of whether Tcf7l2 behaves as an oncogene or tumour suppressor gene in colorectal cancer, its downstream position in the Wnt signalling pathway would make it more attractive than upstream components, for several reasons.

First, targeting an upstream component is more likely to cause unintended side effects compared to downstream components, because of the divergent nature of signalling pathways. Even targeting β -catenin and Tcf7l2, both components that are relatively downstream in the canonical pathway, leads to differential activation of the FoxO signalling pathway. β -catenin knockdown results in accumulation of phosphorylated Foxo4 and hence downregulation of the FoxO pathway²⁹³. On the contrary, Tcf7l2 knockdown results in dephosphorylation of Foxo4 and upregulation of tumour-suppressive FoxO pathway targets²⁹³. This translates into better killing efficiency and chemosensitivity when Tcf7l2, as opposed to β -catenin, is repressed in colorectal cancer cell lines²⁹³.

Second, targeting an upstream component may not have a therapeutic effect if the cancer-driving mutant component is downstream. In colorectal cancer, mutations have been found in many components along the Wnt signalling pathway. However loss-of-function APC mutations are by far the most frequent, occurring in about 50% and 80% of microsatellite instable and microsatellite stable colorectal cancers respectively²⁸⁷.

Targeting components that are upstream of APC, such as Porcupine, Frizzled receptor or R-spondin ligand would not be beneficial in the large majority of colorectal cancers that bear APC mutations. This would not pose a problem if Tcf7l2 were targeted because it is further downstream in the Wnt signalling pathway compared to APC.

Third, functional redundancy among Tcf/Lef proteins may mean less severe side effects in non-target organs. Targeting the Wnt signalling pathway is a double-edged sword; while it may inhibit tumour growth, it is also detrimental to healthy tissues because of its essential role in adult stem cell maintenance. However if Tcf/Lef proteins were co-expressed and could compensate for one another, selectively targeting one member would still leave the other intact to transduce Wnt signal. Low level Wnt signalling might inhibit tumour progression and yet be sufficient for growth, albeit at a slower rate. Functional redundancy among Tcf/Lef proteins is common during embryogenesis. It is less well studied in adults but a prominent example would be the expression of both Tcf7l1 and Tcf7l2 in the hair follicle stem cells and basal epidermal cells of skin¹⁷⁷. The two Tcf/Lef proteins seem to have overlapping functions in the skin because wound healing is attenuated in double knockout but not single knockout skin grafts¹⁷⁷.

In chemotherapy, it is common to utilise two or more therapeutic drugs during treatment. The reason for this is to reduce the chances of resistant cancer cells and minimise the toxic effect to healthy cells. The added benefit of this strategy is that multiple signal transduction pathways are targeted simultaneously so there is a synergistic effect. If Tcf7l2 turns out to be a potential therapeutic target for colorectal cancer, it can be used in combination with other chemotherapy drugs for the best outcome.

4.4.5. Transcription factors as druggable targets

Although there is much therapeutic potential in targeting transcription factors, it is notoriously difficult because transcription factors do not conform to the conventions of druggable targets. Transcription factors lack deep ligand-binding pockets, unlike enzymes and receptors²⁹⁴, making them poor targets of small molecule drugs. Furthermore the potential drug has to cross the plasma and nuclear membrane barriers before reaching the transcription factor in the nucleus. This may pose a problem with large-sized drugs, such as biologics, since the drug cannot readily diffuse across the plasma membrane.

Rather than targeting the recalcitrant transcription factor, it might be more productive to go after proteins that form key interactions with the transcription factor. One such example is the small molecule inhibitor, Nutlin. It abolishes the interaction between MDM2 and the p53 transcription factor, by blocking the p53 binding pocket on MDM2²⁹⁵. This stabilises the p53 tumour suppressor, resulting in growth inhibition of tumour xenografts in mice²⁹⁵. RG7112, an optimised derivative of Nutlin with better efficacy²⁹⁶, has shown promising progress in phase I clinical trials as a treatment for Leukaemia²⁹⁷. Another reason for using an indirect strategy when targeting p53 is because it is a tumour suppressor. Therefore the intention is to prevent the E3 ubiquitin ligase, MDM2, from marking p53 for proteasomal degradation^{298, 299}. If Tcf712 turns out to be a tumour suppressor, a similar strategy could be employed. A possible target would be Nrf1, which is an E3 ubiquitin ligase, that ubiquitinylates Tcf712 and targets it for degradation³⁰⁰.

Some innovative strategies rely on other classes of drugs, besides small molecule inhibitors. Oligonucleotide drugs such as siRNAs or antisense RNAs, block protein expression at the translational level^{301, 302}. Hence they circumvent the difficulties of targeting the protein version of the transcription factor. A fair number of antisense RNAs are in development, and AZD9150 is an example of a promising antisense RNA that targets a transcription factor. It blocks translation of STAT3 thereby inhibiting growth of tumour xenografts in mice³⁰³ and preventing anti-tumour immune suppression³⁰⁴⁻³⁰⁶. STAT3 overexpression has been implicated in many cancers⁶⁶, so it is not surprising that AZD9150 is in phase Ib/II clinical trials as a potential treatment for multiple solid tumours including non-small cell lung carcinoma, hepatocellular carcinoma and metastatic squamous cell carcinoma³⁰⁷⁻³⁰⁹. Another type of oligonucleotide drug that could target transcription factors is aptamers. These are short RNA or DNA oligonucleotides that bind specifically to proteins^{310, 311}. Transcription factors should be amenable to aptamer binding since they interact directly with DNA to perform their functions. Once bound, the aptamer could inhibit a transcription factor's ability to form the protein-DNA and protein-protein interactions that are required for assembly of the transcriptional machinery.

Targeting transcription factors is not futile. It is only a matter of finding alternatives to small molecule drugs. Regardless of the class of drug, it is important to ensure that the drug is specific for its intended target to reduce unwanted side effects. Small molecule drugs are quite promiscuous and can bind anywhere from two to five or more targets³¹². It would be difficult to increase their specificity because it requires intimate knowledge of the interaction between the drug and its target. Conversely, it might be easier to minimise or completely eliminate off-target effects of antisense RNA drugs because they bind their targets by Watson and Crick base pairing and share a universal mechanism of action³¹³.

CHAPTER 5

Conclusions and Future Perspectives

In this thesis, I presented a Tet-inducible system in mice for repression of a target gene. The lynchpin of this system is the trans-silencer, which is a fusion between rTetR and the KRAB domain of Kid-1 protein. DNA-binding specificity is conferred by rTetR while the KRAB domain has transcriptional repression function. In the presence of doxycycline, the trans-silencer binds to the TRE sequence in a dose-dependent manner and represses transcription from nearby promoters. The advantage of this repression system is that the target gene is repressed, rather than overexpressed. Therefore pleiotropic effects that may arise due to overexpression will not be a moot point.

To utilise this system for inducible and reversible repression of a target gene *in vivo*, two transgenic mouse strains are required. First, a general strain that ubiquitously expresses the trans-silencer. Second, a target-specific strain where the TRE sequence is inserted in close proximity to the target gene's promoter. The system is only functional in double transgenic mice; the target gene is repressed when double transgenic mice are given doxycycline.

I have taken a significant step in building this Tet-inducible repression system into a platform for generating inducible and reversible transgenic mice, which can be used to target other genes. Both the general trans-silencer strain and the target-specific TRE strain were created in-house, by using CRISPR/Cas9 gene editing technology. In creating these two transgenic strains, I have resolved teething issues and optimised the transgenesis process. Therefore it will be relatively straightforward to create TRE

strains targeting other genes of interest, by following the optimised protocol described in this thesis. In fact, my colleagues have already created several other TRE strains and used them in combination with the trans-silencer strain for target gene repression. The application of this platform is not restricted to drug target identification and validation. It can also be used for gene function studies to investigate the physiological or pathological relevance of any gene.

Tcf7l2 was selected as the gene of interest because it is a key player in the all-important Wnt signalling pathway. This pathway has an intrinsic role in colorectal cancer so modulating the Wnt signalling pathway activation via *Tcf7l2* could potentially be therapeutic. However targeting Tcf7l2 would be a pointless endeavour, without knowledge as to whether Tcf7l2 is an oncogene or tumour suppressor gene. The role of Tcf7l2 in colorectal cancer has been difficult to resolve because the existing Tcf7l2 transgenic mouse strains are ill equipped for studying a gene that is perinatal lethal. The inducible and reversible Tcf7l2 mouse, which I have created, overcomes the limitations of the previous transgenic mouse strains. Therefore it is a powerful tool to study the exact role of Tcf7l2 in colorectal cancer and possibly other diseases or homeostatic processes.

CHAPTER 6

References

1. Tivol, E.A. et al. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* **3**, 541-547 (1995).
2. Waterhouse, P. et al. Lymphoproliferative Disorders with Early Lethality in Mice Deficient in Ctla-4. *Science* **270**, 985 (1995).
3. Iwai, Y. et al. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proceedings of the National Academy of Sciences* **99**, 12293 (2002).
4. Leach, D.R., Krummel, M.F. & Allison, J.P. Enhancement of Antitumor Immunity by CTLA-4 Blockade. *Science* **271**, 1734-1736 (1996).
5. Hodi, F.S. et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *New England Journal of Medicine* **363**, 711-723 (2010).
6. Topalian, S.L. et al. Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *New England Journal of Medicine* **366**, 2443-2454 (2012).
7. Seidel, J.A., Otsuka, A. & Kabashima, K. Anti-PD-1 and Anti-CTLA-4 Therapies in Cancer: Mechanisms of Action, Efficacy, and Limitations. *Frontiers in oncology* **8**, 86-86 (2018).
8. Bryan-Marrugo, O.L. et al. History and progress of antiviral drugs: From acyclovir to direct-acting antiviral agents (DAAs) for Hepatitis C. *Medicina Universitaria* **17**, 165-174 (2015).
9. Elion, G.B. et al. Selectivity of action of an antiherpetic agent, 9-(2-hydroxyethoxymethyl) guanine. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5716-5720 (1977).

10. Elion, G.B. Mechanism of action and selectivity of acyclovir. *The American Journal of Medicine* **73**, 7-13 (1982).
11. McGuirt, P.V. & Furman, P.A. Acyclovir inhibition of viral DNA chain elongation in herpes simplex virus-infected cells. *The American Journal of Medicine* **73**, 67-71 (1982).
12. O Ohlson, L. et al. Risk factors for Type 2 (non-insulin-dependent) diabetes mellitus. Thirteen and one-half years of follow-up of the participants in a study of Swedish men born in 1913, Vol. 31. (1988).
13. Hamman, R.F. Genetic and environmental determinants of non-insulin-dependent diabetes mellitus (NIDDM). *Diabetes/Metabolism Reviews* **8**, 287-338 (1992).
14. Tuomilehto, J. et al. Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. *New England Journal of Medicine* **344**, 1343-1350 (2001).
15. Knowler, W.C. et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine* **346**, 393-403 (2002).
16. Peters, J.L., Cnudde, F. & Gerats, T. Forward genetics and map-based cloning approaches. *Trends in Plant Science* **8**, 484-491 (2003).
17. Moresco, E.M.Y., Li, X. & Beutler, B. Going forward with genetics: recent technological advances and forward genetics in mice. *The American journal of pathology* **182**, 1462-1473 (2013).
18. Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G. & Jenkins, N.A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221 (2005).
19. Miskey, C., Izsvák, Z., Plasterk, R.H. & Ivics, Z. The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic acids research* **31**, 6873-6881 (2003).
20. Ivics, Z. et al. Transposon-mediated genome manipulation in vertebrates. *Nature methods* **6**, 415-422 (2009).

21. Laathanachareon, T., Tamayo-Ramos, J.A., Nijse, B. & Schaap, P.J. Forward Genetics by Genome Sequencing Uncovers the Central Role of the *Aspergillus niger* *goxB* Locus in Hydrogen Peroxide Induced Glucose Oxidase Expression. *Frontiers in microbiology* **9**, 2269-2269 (2018).
22. Mascher, M. et al. Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biology* **15**, R78 (2014).
23. Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Reviews Genetics* **15**, 662 (2014).
24. van Nimwegen, K.J.M. et al. Is the \$1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clinical Chemistry* **62**, 1458 (2016).
25. Shalem, O., Sanjana, N.E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nature reviews. Genetics* **16**, 299-311 (2015).
26. Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246-1260 (2015).
27. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337** (2012).
28. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109** (2012).
29. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308 (2013).
30. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)* **343**, 84-87 (2014).
31. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)* **343**, 80-84 (2014).
32. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487 (2014).

33. Brown, S.D.M. & Moore, M.W. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Disease models & mechanisms* **5**, 289-292 (2012).
34. Brown, S.D.M. & Moore, M.W. The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. *Mammalian genome : official journal of the International Mammalian Genome Society* **23**, 632-640 (2012).
35. Austin, C.P. et al. The knockout mouse project. *Nature genetics* **36**, 921-924 (2004).
36. Lloyd, K.C.K. A knockout mouse resource for the biomedical research community. *Annals of the New York Academy of Sciences* **1245**, 24-26 (2011).
37. Zhang, X., Wang, W., Xiao, K. & Shi, L. in An Omics Perspective on Cancer Research. (ed. W.C.S. Cho) 235-247 (Springer Netherlands, Dordrecht; 2010).
38. McCarthy, M.I. & Hirschhorn, J.N. Genome-wide association studies: potential next steps on a genetic journey. *Human molecular genetics* **17**, R156-R165 (2008).
39. Gupta, V., Saraswathy, K., Khadgawat, R. & Sachdeva, M. Genome-wide Association: “A Revolutionary Approach”, Vol. 9. (2009).
40. Chen, R. & Jeong, S.S. Functional prediction: identification of protein orthologs and paralogs. *Protein science : a publication of the Protein Society* **9**, 2344-2353 (2000).
41. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586-1597 (2016).
42. Fatehullah, A., Tan, S.H. & Barker, N. Organoids as an in vitro model of human development and disease. *Nature Cell Biology* **18**, 246 (2016).
43. Curtis, C.G., Bilyard, K. & Stephenson, H. Ex Vivo Metrics™, a preclinical tool in new drug development. *Journal of Translational Medicine* **6**, 5-5 (2008).
44. Bain, J. et al. The selectivity of protein kinase inhibitors: a further update. *The Biochemical journal* **408**, 297-315 (2007).
45. Fedorov, O. et al. A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proceedings of the National Academy of Sciences* **104**, 20523 (2007).
46. Milhavet, O., Gary, D.S. & Mattson, M.P. RNA interference in biology and medicine. *Pharmacological reviews* **55**, 629-648 (2003).

47. Houtkooper, R.H. et al. The metabolic footprint of aging in mice. *Scientific reports* **1**, 134 (2011).
48. Tarantino, L. & Bucan, M. Dissection of behavior and psychiatric disorders using the mouse as a model. *Human molecular genetics* **9**, 953-965 (2000).
49. Glazko, G.V. & Nei, M. Estimation of Divergence Times for Major Lineages of Primate Species. *Molecular Biology and Evolution* **20**, 424-434 (2003).
50. Arnason, U. et al. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. *Gene* **421**, 37-51 (2008).
51. Beura, L.K. et al. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* **532**, 512 (2016).
52. Rossi, A. et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230-233 (2015).
53. Zhang, J. et al. Conditional gene manipulation: Cre-ating a new biological era. *Journal of Zhejiang University. Science. B* **13**, 511-524 (2012).
54. Branda, C.S. & Dymecki, S.M. Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. *Developmental cell* **6**, 7-28 (2004).
55. Gu, H., Marth, J.D., Orban, P.C., Mossmann, H. & Rajewsky, K. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103-106 (1994).
56. Lakso, M. et al. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6232-6236 (1992).
57. Semprini, S. et al. Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. *Nucleic acids research* **35**, 1402-1410 (2007).
58. Loonstra, A. et al. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9209-9214 (2001).

59. Gilbert, L.A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154** (2013).
60. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152** (2013).
61. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676 (2014).
62. Cencic, R. et al. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage, Vol. 9. (2014).
63. Yang, M. et al. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell reports* **22**, 372-382 (2018).
64. Yekta, S., Shih, I.h. & Bartel, D.P. MicroRNA-Directed Cleavage of *HOXB8* mRNA. *Science* **304**, 594 (2004).
65. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics* **12**, 99 (2011).
66. Wu, P. et al. Prognostic role of STAT3 in solid tumors: a systematic review and meta-analysis. *Oncotarget* **7**, 19863-19883 (2016).
67. Rehwinkel, J.A.N., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11**, 1640-1647 (2005).
68. Feil, R. et al. Ligand-activated site-specific recombination in mice. *Proceedings of the National Academy of Sciences* **93**, 10887 (1996).
69. Feil, S., Valtcheva, N. & Feil, R. in *Gene Knockout Protocols: Second Edition*. (eds. W. Wurst & R. Kühn) 343-363 (Humana Press, Totowa, NJ; 2009).
70. Parikh, I., Rajendran, K.G., Su, J.L., Lopez, T. & Sar, M. Are estrogen receptors cytoplasmic or nuclear? Some immunocytochemical and biochemical studies. *Journal of steroid biochemistry* **27**, 185-192 (1987).
71. Danielian, P.S., White, R., Hoare, S.A., Fawell, S.E. & Parker, M.G. Identification of residues in the estrogen receptor that confer differential sensitivity to estrogen and hydroxytamoxifen. *Molecular endocrinology* **7**, 232-240 (1993).

72. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5547-5551 (1992).
73. Das, A.T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Current gene therapy* **16**, 156-167 (2016).
74. Baron, U. & Bujard, H. Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. (1997).
75. Costello, A. et al. Leaky Expression of the TET-On System Hinders Control of Endogenous miRNA Abundance. *Biotechnol J*, e1800219 (2018).
76. Roney, I.J., Rudner, A.D., Couture, J.-F. & Kærn, M. Improvement of the reverse tetracycline transactivator by single amino acid substitutions that reduce leaky target gene expression to undetectable levels. *Scientific reports* **6**, 27697 (2016).
77. Blau, J. et al. Three functional classes of transcriptional activation domain. *Molecular and Cellular Biology* **16**, 2044-2055 (1996).
78. Kaneto, H. et al. PDX-1/VP16 Fusion Protein, Together With NeuroD or Ngn3, Markedly Induces Insulin Gene Transcription and Ameliorates Glucose Tolerance. *Diabetes* **54**, 1009 (2005).
79. Zhu, Z., Ma, B., Homer, R.J., Zheng, T. & Elias, J.A. Use of the Tetracycline-controlled Transcriptional Silencer (tTS) to Eliminate Transgene Leak in Inducible Overexpression Transgenic Mice. *Journal of Biological Chemistry* **276**, 25222-25229 (2001).
80. Lamartina, S. et al. Construction of an rtTA2s-m2/ttskid-Based transcription regulatory switch that displays no basal activity, good inducibility, and high responsiveness to doxycycline in mice and Non-Human primates. *Molecular Therapy* **7**, 271-280 (2003).
81. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC biotechnology* **10**, 81 (2010).
82. Gill, G. & Ptashne, M. Negative effect of the transcriptional activator GAL4. *Nature* **334**, 721 (1988).

83. Berger, S.L. et al. Genetic isolation of ADA2: A potential transcriptional adaptor required for function of certain acidic activation domains. *Cell* **70**, 251-265 (1992).
84. Lee, M.-s. et al. Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* **405**, 360 (2000).
85. Otth, C. et al. AbetaPP induces cdk5-dependent tau hyperphosphorylation in transgenic mice Tg2576. *Journal of Alzheimer's disease : JAD* **4**, 417-430 (2002).
86. Crews, L., Patrick, C., Adame, A., Rockenstein, E. & Masliah, E. Modulation of aberrant CDK5 signaling rescues impaired neurogenesis in models of Alzheimer's disease. *Cell Death & Disease* **2**, e120 (2011).
87. Piedrahita, D. et al. Silencing of CDK5 Reduces Neurofibrillary Tangles in Transgenic Alzheimer's Mice. *The Journal of Neuroscience* **30**, 13966 (2010).
88. Patrick, G.N., Zhou, P., Kwon, Y.T., Howley, P.M. & Tsai, L.-H. p35, the Neuronal-specific Activator of Cyclin-dependent Kinase 5 (Cdk5) Is Degraded by the Ubiquitin-Proteasome Pathway. *Journal of Biological Chemistry* **273**, 24057-24064 (1998).
89. Saito, T. et al. Single App knock-in mouse models of Alzheimer's disease. *Nature Neuroscience* **17**, 661 (2014).
90. Saito, T., Matsuba, Y., Yamazaki, N., Hashimoto, S. & Saido, T.C. Calpain Activation in Alzheimer's Model Mice Is an Artifact of APP and Presenilin Overexpression. *The Journal of Neuroscience* **36**, 9933 (2016).
91. Deuschle, U., Meyer, W.K. & Thiesen, H.J. Tetracycline-reversible silencing of eukaryotic promoters. *Molecular and Cellular Biology* **15**, 1907-1914 (1995).
92. Mallo, M., Kanzler, B. & Ohnemus, S. Reversible gene inactivation in the mouse. *Genomics* **81**, 356-360 (2003).
93. Chen, X. et al. Probing the impact of chromatin conformation on genome editing tools. *Nucleic acids research* **44**, 6482-6492 (2016).
94. Gordon, D.J., Resio, B. & Pellman, D. Causes and consequences of aneuploidy in cancer. *Nature reviews. Genetics* **13**, 189-203 (2012).
95. Jaenisch, R., Fan, H. & Croker, B. Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and

- leukemogenesis in the adult animal. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 4008-4012 (1975).
96. Jaenisch, R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 1260-1264 (1976).
 97. Lesbats, P., Engelman, A.N. & Cherepanov, P. Retroviral DNA Integration. *Chemical Reviews* **116**, 12730-12757 (2016).
 98. Kuehn, M.R., Bradley, A., Robertson, E.J. & Evans, M.J. A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 295 (1987).
 99. Doetschman, T. et al. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576 (1987).
 100. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348 (1988).
 101. Thomas, K.R. & Capecchi, M.R. Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**, 847 (1990).
 102. Liu, X. et al. Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Developmental dynamics : an official publication of the American Association of Anatomists* **209**, 85-91 (1997).
 103. Dean, W. et al. Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development* **125**, 2273-2282 (1998).
 104. Ding, Q. et al. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell stem cell* **12**, 393-394 (2013).
 105. Cai, M. & Yang, Y. Targeted genome editing tools for disease modeling and gene therapy. *Current gene therapy* **14**, 2-9 (2014).

106. Gilles, A.F. & Averof, M. Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *EvoDevo* **5**, 1-13 (2014).
107. Barrangou, R. et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315** (2007).
108. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468** (2010).
109. Mohanraju, P. et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353** (2016).
110. Hale, C.R. et al. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945-956 (2009).
111. Mougiakos, I., Bosma, E.F., de Vos, W.M., van Kranenburg, R. & van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends in biotechnology* **34**, 575-587 (2016).
112. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
113. Hale, C., Kleppe, K., Terns, R.M. & Terns, M.P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *Rna* **14**, 2572-2579 (2008).
114. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).
115. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505-510 (2015).
116. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
117. Mali, P. et al. RNA-Guided human genome engineering via Cas9. *Science* **339** (2013).
118. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918 (2013).
119. Yang, H. et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-1379 (2013).

120. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014).
121. Garside, E.L. & MacMillan, A.M. Cas9 in close-up. *Nature biotechnology* **32**, 338-340 (2014).
122. Horii, T. et al. Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Scientific reports* **4**, 4513 (2014).
123. Mashiko, D. et al. Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Scientific reports* **3**, 3355 (2013).
124. Li, D. et al. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nature biotechnology* **31**, 681-683 (2013).
125. Weber, J.L. Analysis of sequences from the extremely A + T-rich genome of *Plasmodium falciparum*. *Gene* **52**, 103-109 (1987).
126. Zetsche, B. et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759-771 (2015).
127. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).
128. Clevers, H. & Nusse, R. Wnt/beta-catenin signaling and disease. *Cell* **149**, 1192-1205 (2012).
129. Dajani, R. et al. Structural basis for recruitment of glycogen synthase kinase 3 β to the axin—APC scaffold complex. *The EMBO Journal* **22**, 494 (2003).
130. Amit, S. et al. Axin-mediated CKI phosphorylation of β -catenin at Ser 45: a molecular switch for the Wnt pathway. *Genes & development* **16**, 1066-1076 (2002).
131. Su, Y. et al. APC Is Essential for Targeting Phosphorylated β -Catenin to the SCF β -TrCP Ubiquitin Ligase. *Molecular cell* **32**, 652-661 (2008).
132. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. β -catenin is a target for the ubiquitin—proteasome pathway. *The EMBO Journal* **16**, 3797 (1997).
133. Hsieh, J.-C. et al. A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* **398**, 431 (1999).

134. Dennis, S., Aikawa, M., Szeto, W., Amore, P.A. & Papkoff, J. A secreted frizzled related protein, FrzA, selectively associates with Wnt-1 protein and regulates wnt-1 signaling. *Journal of cell science* **112**, 3815 (1999).
135. Glinka, A. et al. Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357 (1998).
136. Mao, B. et al. Kremen proteins are Dickkopf receptors that regulate Wnt/ β -catenin signalling. *Nature* **417**, 664 (2002).
137. Nusse, R. & Clevers, H. Wnt/ β -Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999 (2017).
138. Jho, E.-h. et al. Wnt/ β -Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway. *Molecular and Cellular Biology* **22**, 1172 (2002).
139. Yan, D. et al. Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14973-14978 (2001).
140. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003 (2007).
141. Hao, H.-X. et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195 (2012).
142. Koo, B.-K. et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* **488**, 665 (2012).
143. Kuhnert, F. et al. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proceedings of the National Academy of Sciences* **101**, 266 (2004).
144. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes & development* **17**, 1709-1713 (2003).

145. Ireland, H. et al. Inducible cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of β -catenin. *Gastroenterology* **126**, 1236-1246 (2004).
146. Luongo, C., Moser, A.R., Gledhill, S. & Dove, W.F. Loss of Δ Apc⁺ in Intestinal Adenomas from Min Mice. *Cancer Research* **54**, 5947 (1994).
147. Moser, A.R., Pitot, H.C. & Dove, W.F. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* **247**, 322 (1990).
148. Batlle, E. et al. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. (2002).
149. Drescher, U. et al. In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands for Eph receptor tyrosine kinases. *Cell* **82**, 359-370 (1995).
150. Hrckulak, D., Kolar, M., Strnad, H. & Korinek, V. TCF/LEF Transcription Factors: An Update from the Internet Resources. *Cancers* **8** (2016).
151. Korinek, V. et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. (1998).
152. Savic, D. et al. Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. (2011).
153. van Es, J.H. et al. A Critical Role for the Wnt Effector Tcf4 in Adult Intestinal Homeostatic Self-Renewal. *Molecular and Cellular Biology* **32**, 1918-1927 (2012).
154. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186-193 (2004).
155. Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626-638 (2005).
156. Ferlay, J. et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**, 2893-2917 (2010).
157. O'Connell, M.R. et al. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263-266 (2014).

158. Riihimäki, M., Hemminki, A., Sundquist, J. & Hemminki, K. Patterns of metastasis in colon and rectal cancer. *Scientific reports* **6**, 29765 (2016).
159. Weiss, L. et al. Haematogenous metastatic patterns in colonic carcinoma: An analysis of 1541 necropsies. *The Journal of pathology* **150**, 195-203 (1986).
160. Rao, T.P. & Kuhl, M. An updated overview on Wnt signaling pathways: a prelude for more. *Circulation research* **106**, 1798-1806 (2010).
161. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).
162. Albuquerque, C. et al. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the β -catenin signaling cascade. *Human molecular genetics* **11**, 1549-1560 (2002).
163. Christie, M. et al. Different APC genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/ β -catenin signalling thresholds for tumourigenesis. *Oncogene* **32**, 4675 (2012).
164. Korinek, V. Constitutive Transcriptional Activation by a beta -Catenin-Tcf Complex in APC-/- Colon Carcinoma. *Science* **275**, 1784-1787 (1997).
165. Dow, Lukas E. et al. Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. *Cell* **161**, 1539-1552 (2015).
166. Liu, J. et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proceedings of the National Academy of Sciences* **110**, 20224 (2013).
167. Madan, B. et al. Wnt addiction of genetically defined cancers reversed by PORCN inhibition. *Oncogene* **35**, 2197-2207 (2016).
168. Gurney, A. et al. Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors. *Proceedings of the National Academy of Sciences* **109**, 11717 (2012).
169. Sherwood, V. WNT Signaling: an Emerging Mediator of Cancer Cell Metabolism? *Molecular and Cellular Biology* **35**, 2 (2015).
170. Varelas, X. et al. The Hippo Pathway Regulates Wnt/ β -Catenin Signaling. *Developmental cell* **18**, 579-591 (2010).

171. Ding, M. & Wang, X. Antagonism between Hedgehog and Wnt signaling pathways regulates tumorigenicity. *Oncology Letters* **14**, 6327-6333 (2017).
172. Tang, W. et al. A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9697-9702 (2008).
173. Angus-Hill, M.L., Elbert, K., Hidalgo, J. & Capecchi, M.R. T-cell factor 4 functions as a tumor suppressor whose disruption modulates colon cell proliferation and tumorigenesis. (2011).
174. da Silva Xavier, G. et al. Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. *Diabetologia* **55**, 2667-2676 (2012).
175. Institute, W.T.S. Alleles produced for the EUCOMM and EUCOMMTools projects by the Wellcome Trust Sanger Institute. *MGI Direct Data Submission* (2010).
176. Arango, D. et al. Villin Expression Is Frequently Lost in Poorly Differentiated Colon Cancer. *The American journal of pathology* **180**, 1509-1521 (2012).
177. Nguyen, H. et al. Tcf3 and Tcf4 are essential for long-term homeostasis of skin epithelia. *Nature genetics* **41**, 1068-1075 (2009).
178. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research* **42**, W401-W407 (2014).
179. Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic acids research* **44**, W272-W276 (2016).
180. Gossen, M. et al. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766 (1995).
181. Witzgall, R., O'Leary, E., Leaf, A., Onaldi, D. & Bonventre, J.V. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 4514-4518 (1994).

182. Auerbach, W. et al. Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines. (2000).
183. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671 (2012).
184. Brinkman, E.K. et al. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Molecular cell* **70**, 801-813.e806 (2018).
185. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes & development* **5**, 1513-1523 (1991).
186. Zambrowicz, B.P. et al. Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β -galactosidase in mouse embryos and hematopoietic cells. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3789-3794 (1997).
187. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507** (2014).
188. Xu, H. et al. Sequence determinants of improved CRISPR sgRNA design. *Genome research* **25**, 1147-1157 (2015).
189. Niwa, O., Yokota, Y., Ishida, H. & Sugahara, T. Independent mechanisms involved in suppression of the moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell* **32**, 1105-1113 (1983).
190. Chen, C.-m., Krohn, J., Bhattacharya, S. & Davies, B. A Comparison of Exogenous Promoter Activity at the ROSA26 Locus Using a PhiC31 Integrase Mediated Cassette Exchange Approach in Mouse ES Cells. *PloS one* **6**, e23376 (2011).
191. Nyabi, O. et al. Efficient mouse transgenesis using Gateway-compatible ROSA26 locus targeting vectors and F1 hybrid ES cells. *Nucleic acids research* **37**, e55-e55 (2009).
192. Groner, A.C. et al. KRAB-Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. *PLOS Genetics* **6**, e1000869 (2010).

193. Hasty, P., Rivera-Pérez, J., Chang, C. & Bradley, A. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Molecular and Cellular Biology* **11**, 4509-4517 (1991).
194. Antonchuk, J., Sauvageau, G. & Humphries, R.K. HOXB4-Induced Expansion of Adult Hematopoietic Stem Cells Ex Vivo. *Cell* **109**, 39-45 (2002).
195. Pineault, N., Abramovich C Fau - Ohta, H., Ohta H Fau - Humphries, R.K. & Humphries, R.K. Differential and common leukemogenic potentials of multiple NUP98-Hox fusion proteins alone or with Meis1. (2004).
196. Takiguchi, M. et al. Variability of Inducible Expression across the Hematopoietic System of Tetracycline Transactivator Transgenic Mice. *PloS one* **8**, e54009 (2013).
197. Holmberg, J. et al. EphB Receptors Coordinate Migration and Proliferation in the Intestinal Stem Cell Niche. *Cell* **125**, 1151-1163 (2006).
198. Germann, M. et al. Tripartite interactions between Wnt signaling, Notch and Myb for stem/progenitor cell functions during intestinal tumorigenesis. *Stem Cell Research* **13**, 355-366 (2014).
199. Langlands, A.J., Carroll, T.D., Chen, Y. & Näthke, I. Chir99021 and Valproic acid reduce the proliferative advantage of Apc mutant cells. *Cell Death & Disease* **9**, 255 (2018).
200. Pengue, G., Calabrò, V., Bartoli, P.C., Pagliuca, A. & Lania, L. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic acids research* **22**, 2908-2914 (1994).
201. Bellí, G., Garí, E., Piedrafita, L., Aldea, M. & Herrero, E. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic acids research* **26**, 942-947 (1998).
202. Milch, R.A., Tobie, J.E., Robinson, R.A. & Evans, C.B. A MICROSCOPIC STUDY OF TETRACYCLINE LOCALIZATION IN SKELETAL NEOPLASMS. *Journal of Histochemistry & Cytochemistry* **9**, 261-270 (1961).

203. Milch Ra Fau - Rall, D.P., Rall Dp Fau - Tobie, J.E. & Tobie, J.E. Bone localization of the tetracyclines. (1957).
204. Finerman, G.A.M. & Milch, R.A. In vitro Binding of Tetracyclines to Calcium. *Nature* **198**, 486 (1963).
205. Katlam, S., Deshmukh, Y.A. & Jadhav, P.R. Comparative study of oxytetracycline and doxycycline on calcium chelation: in-vitro assay. *International Journal of Basic & Clinical Pharmacology; Vol 6, No 5 (2017): May 2017* (2017).
206. Barza, M., Brown, R.B., Shanks, C., Gamble, C. & Weinstein, L. Relation between lipophilicity and pharmacological behavior of minocycline, doxycycline, tetracycline, and oxytetracycline in dogs. (1975).
207. Agwuh, K.N. & MacGowan, A. Pharmacokinetics and pharmacodynamics of the tetracyclines including glycylcyclines. (2006).
208. Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biology* **4**, 231-231 (2003).
209. Ying, Y. et al. The Krüppel-associated box repressor domain induces reversible and irreversible regulation of endogenous mouse genes by mediating different chromatin states. *Nucleic acids research* **43**, 1549-1561 (2015).
210. Wiznerowicz, M. et al. The Krüppel-associated Box Repressor Domain Can Trigger de Novo Promoter Methylation during Mouse Early Embryogenesis. *Journal of Biological Chemistry* **282**, 34535-34541 (2007).
211. Lanza, R. et al. Essentials of Stem Cell Biology. (Elsevier Science, 2009).
212. Uusi-Mäkelä, M.I.E. et al. Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (*Danio rerio*). *PloS one* **13**, e0196238 (2018).
213. Horlbeck, M.A. et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, e12677 (2016).
214. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31** (2013).

215. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research* **42**, e168 (2014).
216. Hawkins, R.D. et al. Distinct Epigenomic Landscapes of Pluripotent and Lineage-committed Human Cells. *Cell stem cell* **6**, 479-491 (2010).
217. Efroni, S. et al. Global transcription in pluripotent embryonic stem cells. *Cell stem cell* **2**, 437-447 (2008).
218. Krejčí, J. et al. Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *Journal of Cellular Physiology* **219**, 677-687 (2009).
219. Kass, E.M. et al. Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5564-5569 (2013).
220. Tichy, E.D. et al. Mouse Embryonic Stem Cells, but Not Somatic Cells, Predominantly Use Homologous Recombination to Repair Double-Strand DNA Breaks. *Stem Cells and Development* **19**, 1699-1711 (2010).
221. Veres, A. et al. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell stem cell* **15**, 27-30 (2014).
222. Smith, C. et al. Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. *Cell stem cell* **15**, 12-13 (2014).
223. Tsai, S.Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology* **32** (2014).
224. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods* **12**, 237 (2015).
225. Frock, R.L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature biotechnology* **33**, 179-186 (2015).
226. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31** (2013).

227. Ran, F.A. et al. Double nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154** (2013).
228. Guilinger, J.P., Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. (2014).
229. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. (2014).
230. Ramakrishna, S. et al. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. (2014).
231. Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187 (2014).
232. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* **32**, 279-284 (2014).
233. Zhang, J.-P. et al. Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency. *Scientific reports* **6**, 28566 (2016).
234. Pogson, M., Parola, C., Kelton, W.J., Heuberger, P. & Reddy, S.T. Immunogenomic engineering of a plug-and-(dis)play hybridoma platform. *Nature communications* **7**, 12535 (2016).
235. Chu, V.T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548 (2015).
236. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotech* **33**, 538-542 (2015).
237. Lorenz, E., Congdon, C. & Uphoff, D. Modification of Acute Irradiation Injury in Mice and Guinea-Pigs by Bone Marrow Injections. *Radiology* **58**, 863-877 (1952).
238. Congdon, C.C., Uphoff, D. & Lorenz, E. Modification of acute irradiation injury in mice and guinea pigs by injection of bone marrow; a histopathologic study. *Journal of the National Cancer Institute* **13**, 73-107 (1952).

239. Kurnick, N.B. & Nokay, N. Repopulation of Bone Marrow in Mice: Number and Type of Cells Required for Post-X-Irradiation Protection. *Radiation Research* **25**, 53-67 (1965).
240. Ruedl, C., Khameneh, H.J. & Karjalainen, K. Manipulation of immune system via immortal bone marrow stem cells. *International Immunology* **20**, 1211-1218 (2008).
241. Von Wittenau, M.S., Twomey, T.M. & Swindell, A.C. The Disposition of Doxycycline by the Rat. *Chemotherapy* **17**, 26-39 (1972).
242. Pindell, M.H., Cull, K.M., Doran, K.M. & Dickison, H.L. ABSORPTION AND EXCRETION STUDIES ON TETRACYCLINE. *Journal of Pharmacology and Experimental Therapeutics* **125**, 287 (1959).
243. Von Wittenau, M.S. & Yeary, R. The excretion and distribution in body fluids of tetracyclines after intravenous administration to dogs. *Journal of Pharmacology and Experimental Therapeutics* **140**, 258 (1963).
244. Mahon, W.A., Wittenberg, J.V.P. & Tuffnel, P.G. Studies on the Absorption and Distribution of Doxycycline in Normal Patients and in Patients with Severely Impaired Renal Function. *Canadian Medical Association Journal* **103**, 1031-1034 (1970).
245. Dotevall, L. & Hagberg, L. Penetration of doxycycline into cerebrospinal fluid in patients treated for suspected Lyme neuroborreliosis. *Antimicrobial Agents and Chemotherapy* **33**, 1078-1080 (1989).
246. Yim Cw Fau - Flynn, N.M., Flynn Nm Fau - Fitzgerald, F.T. & Fitzgerald, F.T. Penetration of oral doxycycline into the cerebrospinal fluid of patients with latent or neurosyphilis. (1985).
247. Corish, P. & Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering, Design and Selection* **12**, 1035-1040 (1999).
248. Sacchetti, A., El Sewedy, T., Nasr, A.F. & Alberti, S. Efficient GFP mutations profoundly affect mRNA transcription and translation rates. *FEBS Letters* **492**, 151-155 (2001).

249. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E.K. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences* **103**, 13004 (2006).
250. Thompson, J.F., Hayes, L.S. & Lloyd, D.B. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* **103**, 171-177 (1991).
251. Li, X. et al. Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *Journal of Biological Chemistry* **273**, 34970-34975 (1998).
252. Mateus, C. & Avery Simon, V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313-1323 (2000).
253. Kitamura, T. et al. Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Experimental Hematology* **31**, 1007-1014 (2003).
254. Rival-Gervier, S. et al. Kinetics and Epigenetics of Retroviral Silencing in Mouse Embryonic Stem Cells Defined by Deletion of the D4Z4 Element. *Molecular Therapy* **21**, 1536-1550 (2013).
255. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. & Rauscher, F.J., 3rd SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & development* **16**, 919-932 (2002).
256. Rowe, H.M. et al. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519-529 (2013).
257. Welling, P.G., Koch, P.A., Lau, C.C. & Craig, W.A. Bioavailability of tetracycline and doxycycline in fasted and nonfasted subjects. (1977).
258. Pfizer VIBRAMYCIN® (doxycycline hyclate) - Material Safety Data Sheet. (2007).
259. Walker, M.K. et al. A less stressful alternative to oral gavage for pharmacological and toxicological studies in mice. *Toxicology and Applied Pharmacology* **260**, 65-69 (2012).

260. Cawthorne, C., Swindell, R., Stratford, I.J., Dive, C. & Welman, A. Comparison of Doxycycline Delivery Methods for Tet-Inducible Gene Expression in a Subcutaneous Xenograft Model. *Journal of Biomolecular Techniques : JBT* **18**, 120-123 (2007).
261. Baron, U. & Bujard, H. in *Methods in Enzymology*, Vol. 327. (eds. J. Thorner, S.D. Emr & J.N. Abelson) 401-421 (Academic Press, 2000).
262. Zhou, X., Vink, M., Klaver, B., Berkhout, B. & Das, A.T. Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Therapy* **13**, 1382 (2006).
263. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
264. Zhao, J., Schug, J., Li, M., Kaestner, K.H. & Grant, S.F.A. Disease-associated loci are significantly over-represented among genes bound by transcription factor 7-like 2 (TCF7L2) in vivo. *Diabetologia* **53**, 2340-2346 (2010).
265. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40**, 638-645 (2008).
266. Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-1345 (2007).
267. Villareal, D.T. et al. TCF7L2 Variant rs7903146 Affects the Risk of Type 2 Diabetes by Modulating Incretin Action. *Diabetes* **59**, 479-485 (2010).
268. Donath, M.Y., Gross, D.J., Cerasi, E. & Kaiser, N. Hyperglycemia-induced beta-cell apoptosis in pancreatic islets of Psammomys obesus during development of diabetes. *Diabetes* **48**, 738 (1999).
269. Maedler, K. et al. Glucose Induces β -Cell Apoptosis Via Upregulation of the Fas Receptor in Human Islets. *Diabetes* **50**, 1683 (2001).
270. Lyssenko, V. et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *Journal of Clinical Investigation* **117**, 2155-2163 (2007).

271. Zhou, Y. et al. TCF7L2 is a master regulator of insulin production and processing. *Human molecular genetics* **23**, 6419-6431 (2014).
272. Loder, Merewyn K., Xavier, Gabriela da S., McDonald, A. & Rutter, Guy A. TCF7L2 controls insulin gene expression and insulin secretion in mature pancreatic β -cells. *Biochemical Society Transactions* **36**, 357 (2008).
273. Weise, A. et al. Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets. *Nucleic acids research* **38**, 1964-1981 (2010).
274. Osmark, P. et al. Unique splicing pattern of the TCF7L2 gene in human pancreatic islets. *Diabetologia* **52**, 850-854 (2009).
275. Le Bacquer, O. et al. TCF7L2 splice variants have distinct effects on β -cell turnover and function. *Human molecular genetics* **20**, 1906-1915 (2011).
276. van de Wetering, M. et al. The β -Catenin/TCF-4 Complex Imposes a Crypt Progenitor Phenotype on Colorectal Cancer Cells. *Cell* **111**, 241-250 (2002).
277. van der Flier, L.G. et al. Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. *Cell* **136**, 903-912 (2009).
278. Yamazaki, M. et al. Intensive Immunofluorescence Staining Methods for Low Expression Protein: Detection of Intestinal Stem Cell Marker LGR5. *Acta Histochemica et Cytochemica* **48**, 159-164 (2015).
279. Wang, L. et al. Methods to determine intestinal permeability and bacterial translocation during liver disease. *Journal of immunological methods* **421**, 44-53 (2015).
280. Brinkmeier, M.L. et al. TCF and Groucho-Related Genes Influence Pituitary Growth and Development. *Molecular endocrinology* **17**, 2152-2161 (2003).
281. Kennell, J.A., O'Leary, E.E., Gummow, B.M., Hammer, G.D. & MacDougald, O.A. T-cell factor 4N (TCF-4N), a novel isoform of mouse TCF-4, synergizes with beta-catenin to coactivate C/EBPalpha and steroidogenic factor 1 transcription factors. *Mol Cell Biol* **23**, 5366-5375 (2003).
282. Barker, N. et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608 (2008).

283. Shapiro, M. et al. Wnt Pathway Activation Predicts Increased Risk of Tumor Recurrence in Patients with Stage I Non-Small Cell Lung Cancer. *Annals of surgery* **257**, 548-554 (2013).
284. Zhu, R. et al. Ascl2 knockdown results in tumor growth arrest by miRNA-302b-related inhibition of colon cancer progenitor cells. *PloS one* **7**, e32170 (2012).
285. Chang, Y.-W. et al. Diverse Targets of β -Catenin during the Epithelial–Mesenchymal Transition Define Cancer Stem Cells and Predict Disease Relapse. *Cancer Research* **75**, 3398 (2015).
286. Drost, J. et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43 (2015).
287. The Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330 (2012).
288. Su, L.K. et al. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668 (1992).
289. Cao, H. et al. The secondary bile acid, deoxycholate accelerates intestinal adenoma–adenocarcinoma sequence in Apc min/+ mice through enhancing Wnt signaling. *Familial Cancer* **13**, 563-571 (2014).
290. Tanaka, T. et al. Dextran sodium sulfate strongly promotes colorectal carcinogenesis in ApcMin/+ mice: Inflammatory stimuli by dextran sodium sulfate results in development of multiple colonic neoplasms. *International Journal of Cancer* **118**, 25-34 (2005).
291. Fischer, M.M. et al. RSPO3 antagonism inhibits growth and tumorigenicity in colorectal tumors harboring common Wnt pathway mutations. *Scientific reports* **7**, 15270 (2017).
292. Osawa, Y. et al. Inhibition of Cyclic Adenosine Monophosphate (cAMP)-response Element-binding Protein (CREB)-binding Protein (CBP)/ β -Catenin Reduces Liver Fibrosis in Mice. *EBioMedicine* **2**, 1751-1758 (2015).
293. Xie, J. et al. Inhibition of Tcf-4 Induces Apoptosis and Enhances Chemosensitivity of Colon Cancer Cells. *PloS one* **7**, e45617 (2012).
294. Dunlap, N.K. & Hurn, D.M. Medicinal Chemistry. (CRC Press, 2018).

295. Vassilev, L.T. et al. In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science* **303**, 844 (2004).
296. Vu, B. et al. Discovery of RG7112: A Small-Molecule MDM2 Inhibitor in Clinical Development. *ACS Medicinal Chemistry Letters* **4**, 466-469 (2013).
297. Andreeff, M. et al. Results of the Phase I Trial of RG7112, a Small-Molecule MDM2 Antagonist in Leukemia. *Clinical Cancer Research* **22**, 868 (2016).
298. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296 (1997).
299. Honda, R., Tanaka, H. & Yasuda, H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett* **420**, 25-27 (1997).
300. Yamada, M. et al. NARF, an Nemo-like Kinase (NLK)-associated Ring Finger Protein Regulates the Ubiquitylation and Degradation of T Cell Factor/Lymphoid Enhancer Factor (TCF/LEF). *Journal of Biological Chemistry* **281**, 20749-20760 (2006).
301. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391** (1998).
302. Hamilton, A.J. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286** (1999).
303. Hong, D. et al. AZD9150, a Next-Generation Antisense Oligonucleotide Inhibitor of STAT3 with Early Evidence of Clinical Activity in Lymphoma and Lung Cancer. *Science translational medicine* **7**, 314ra185-314ra185 (2015).
304. Wang, M.-C., Dolphin, A. & Kitmitto, A. L-type voltage-gated calcium channels: understanding function through structure. *FEBS Letters* **564**, 245-250 (2004).
305. Williams, L., Bradley, L., Smith, A. & Foxwell, B. Signal Transducer and Activator of Transcription 3 Is the Dominant Mediator of the Anti-Inflammatory Effects of IL-10 in Human Macrophages. *The Journal of Immunology* **172**, 567 (2004).
306. Kortylewski, M. et al. Inhibiting Stat3 signaling in the hematopoietic system elicits multicomponent antitumor immunity. *Nature medicine* **11**, 1314 (2005).
307. AZD9150 Plus Durvalumab Alone or in Combination With Chemotherapy in Patients With Advanced, Solid Tumours and in Patients With Non-Small-Cell Lung Cancer.

- (2018) Retrieved from <https://clinicaltrials.gov/ct2> (Identification No. NCT03421353).
308. A Phase I/Ib Study of AZD9150 (ISIS-STAT3Rx) in Patients With Advanced/Metastatic Hepatocellular Carcinoma. (2013) Retrieved from <https://clinicaltrials.gov/ct2> (Identification No. NCT01839604).
 309. Study to Assess MEDI4736 With Either AZD9150 or AZD5069 in Advanced Solid Tumors & Relapsed Metastatic Squamous Cell Carcinoma of Head & Neck. (2018) Retrieved from <https://clinicaltrials.gov/ct2> (Identification No. NCT02499328).
 310. Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **249**, 505 (1990).
 311. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818 (1990).
 312. Hu, Y., Gupta-Ostermann, D. & Bajorath, J. EXPLORING COMPOUND PROMISCUITY PATTERNS AND MULTI-TARGET ACTIVITY SPACES. *Computational and Structural Biotechnology Journal* **9**, e201401003 (2014).
 313. Crooke, S.T. Antisense Drug Technology: Principles, Strategies, and Applications, Second Edition. (Taylor & Francis, 2007).
 314. Takeuchi, R., Choi, M. & Stoddard, B.L. Engineering of customized meganucleases via in vitro compartmentalization and in cellulo optimization. *Methods in molecular biology (Clifton, N.J.)* **1239**, 105-132 (2015).
 315. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39** (2011).
 316. Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nature Reviews Molecular Cell Biology* **15**, 19 (2013).
 317. Wang, Y. et al. Bioengineered Systems and Designer Matrices That Recapitulate the Intestinal Stem Cell Niche, Vol. 5. (2018).
 318. Murphy, M.M., Lawson Ja Fau - Mathew, S.J., Mathew Sj Fau - Hutcheson, D.A., Hutcheson Da Fau - Kardon, G. & Kardon, G. Satellite cells, connective tissue

fibroblasts and their interactions are crucial for muscle regeneration. *Development* (2011).

- 319. Mathew, S.J. et al. Connective tissue fibroblasts and Tcf4 regulate myogenesis. *Development* (2011).
- 320. Skarnes, W.C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342 (2011).

Appendix A

Supplementary Figures

Figure A2.2.1. Timeline for retroviral transduction of bone marrow cells.

Three separate components were prepared concurrently and combined for successful retroviral transduction. Platinum-E cells were used to package retroviral vectors for infection of bone marrow cells. Recipient mice were treated with 5-fluorouracil to deplete differentiated hematopoietic cells, thus leaving the progenitor cells. Tissue culture plates were coated with fibronectin to enhance capture of viral particles onto the plate, prior to spin infection.

Figure A2.3.1. Purified recombinant R-spondin1 resolved by SDS-PAGE.

R-spondin1 was eluted from the Protein A agarose column in 2 ml fractions and analysed by denaturing SDS-PAGE. For fractions 5 to 7, 2.5 µg of total protein was applied to the gel. For fractions 4 and 8, as much protein as possible was applied. 5 µg of Calf Ig was also loaded as a control to ensure that SDS-PAGE was successful. Stacking and resolving gels were 5% and 8% respectively. Bands were visualised by staining with Coomassie Blue stain. The manufacturer estimates that the recombinant R-spondin1 is 60 kDa. This information is based on the protein gel photograph provided in their user manual.

	# 1	# 2	# 3	# 4
Gender:	Male	Male	Male	Female

Figure A3.1.5. ROSA26-rTetR-KRAB(Kid1) chimeras generated with embryonic stem cell clone 16.

Mouse blastocysts were injected with ROSA26-rTetR-KRAB(Kid1) embryonic stem (ES) cells (clone 16), implanted into pseudo-pregnant females and carried to full term. The resulting pups were chimeras because they are derived from two sources: the injected ES cells (agouti coat colour) and host blastocyst (black coat colour).

Figure A3.1.6. Tcf7l2-TRE chimeras generated with embryonic stem cell clone 18.

Mouse blastocysts were injected with Tcf7l2-TRE embryonic stem (ES) cells (clone 18), implanted into pseudo-pregnant females and carried to full term. The resulting pups were chimeras because they derived from two sources: the injected ES cells (agouti coat colour) and host blastocyst (black coat colour). Chimera 4 was considered to be a hermaphrodite because it had reproductive organs of male and female genders. The external genitalia were distinctively male but the internal genitalia appeared to be both female (oviduct on mouse's right side) and male (grossly enlarged testis on mouse's left side).

Figure A3.2.1. Purity of GFP⁺ reporter bone marrow cells after fluorescence-activated cell sorting (FACS).

After FACS, wildtype and heterozygous reporter bone marrow cells were expanded in culture before checking for purity by flow cytometry. First, cells were identified in a forward/side scatter plot (FSC-H/SSC-H). Second, GFP⁺ cells were gated on. Percentages of cells encompassed within the gates are given as pink-coloured numbers.

Figure A3.2.2. Gating strategy for the major hematopoietic cell populations in blood and spleen.

(A–B) Total peripheral blood mononuclear cells (PBMCs) were isolated from blood by Ficoll gradient, stained with antibody cocktail and acquired by flow cytometry. Gates were set using fluorescence minus one (FMO) controls. Data presented are from an untreated bone marrow chimera. (A) Small and large cells, as identified in a forward/side scatter plot (FSC-A/SSC-A), were gated separately. (B) Histogram plots showing the percentage of GFP⁺ cells in the respective blood cell populations. (C–D) Total spleen cells were isolated, stained with antibody cocktail and acquired by flow cytometry. Gates were set using fluorescence minus one (FMO) controls. Data presented are from an untreated bone marrow chimera and a normal C57BL/6 mouse (non-chimera). (C) Since the bone marrow chimeras were generated using immortalised donor cells, which had been in culture for 6 to 8 weeks, the hematopoietic profile differed slightly from the C57BL/6 control. However the presence of all the major splenic cell populations in the C57BL/6 control suggest that the staining was performed properly. (D) Histogram plots showing the percentage of GFP⁺ cells in the respective spleen cell populations from the bone marrow chimera. The neutrophils and monocyte populations were absent. Percentages of cells encompassed within the gates are given as pink-coloured numbers.

Appendix B

Oligonucleotide Sequences and PCR Conditions

Table B1: Oligo pairs that were annealed to obtain the sgRNA for cloning.

Gene	Name	Sequence (5' - NNN -3') ^a
Tcf7l2	sg T1 fwd	<u>CACC</u> G TGCGTCTTGGGCTTTCCCC
	sg T1 rev	<u>AAAC</u> GGGGAAAGCCCAAGACGCAC
Tcf7l2	sg T2 fwd	<u>CACC</u> G CAGCTAAGGGGTAAACCTG
	sg T2 rev	<u>AAAC</u> CAGGTAAACCCCTTAGCTGC
Tcf7l2	sg T3 fwd	<u>CACC</u> G GAGGCCCGGCTGGGCCGCG
	sg T3 rev	<u>AAAC</u> CGCGGCCCAAGCCGGGCCTCC
Tcf7l2	sg T4 fwd	<u>CACC</u> G TAGGGCCCGCGCTTTCCGG
	sg T4 rev	<u>AAAC</u> CCGGAAAGCGCGGGCCCTAC
Tcf7l2	sg T5 fwd	<u>CACC</u> G GATCGGCCACCCAACCTTG
	sg T5 rev	<u>AAAC</u> CAAAGTTGGGTGGCCGATCC
ROSA26	sg R1 fwd	<u>CACC</u> G ACTGGAGTTGCAGATCACG
	sg R1 rev	<u>AAAC</u> CGTGATCTGCAACTCCAGTC

All oligonucleotides were modified with a 5'-Phosphate group to facilitate cloning.

^a The underlined sequence indicate additional bases that correspond to the 5' overhangs when the oligo pair was annealed. This was to create complementary sticky ends to Bbs1 sticky ends for direct cloning into the pSpCas9(Chimeric_BB)-2A-GFP vector.

Annealing thermocycling program:

95°C	5 min
94 to 4°C	1 min (-1°C with each cycle)
16°C	Hold

Table B2: Primer pairs for screening embryonic stem (ES) cell clones.

Primers	Sequence (5' - NNN -3') ^a	DNA Pol	Anneal Temp (°C)	Extension time
SR1 + SR2	CTCGTCGCTGATTGGCTTC TACTCCACCCATTGACGTC	KapaTaq	65	15 s
SR3 + SR4	GTGATGCTGGAGAACTACAG GTAATGGCTGCAGACTTAGC	GoTaq	61	2 min 20 s
ST1 + ST2	TCCCGGGAAAGTTTGGAAG CGACAATTCGCGAATCGATG	KapaTaq	60	40 s
ST3 + ST4	CATCGATTCGCGAATTGTCG GAAGGAGGTCTGCGTTCTAG	GoTaq	55	1 min 40 s
ST1 + ST4	TCCCGGGAAAGTTTGGAAG GAAGGAGGTCTGCGTTCTAG	KapaTaq	60	3 min

All KapaTaq reactions were carried out with the provided GC buffer, recommended for amplicons with high GC content.

All PCR reactions were performed for 35 cycles, and using the thermocycling parameters specified by the manufacturer.

Table B3: Primer pairs for genotyping transgenic mice by multiplex PCR.

Primers	Sequence (5' - NNN -3') ^a	DNA Pol	Anneal Temp (°C)	Extension time
GR1 +	ACGTTTCCGACTTGAGTTG	GoTaq	58	1 min
GR2 +	CCTGTGTCAGCAAGGCTTC			
GR3	CTTTAAGCCTGCCCAGAAG			
GT1 +	CGAATGTCTGCAGTGAGTG	GoTaq	58	1 min
GT2 +	CATCGATTCGCGAATTGTCG			
GT3	CTGTGCGTCTTGGGCTTTC			

All PCR reactions were performed for 35 cycles, and using the thermocycling parameters specified by the manufacturer.

Table B4: Primer pairs for end-point, semi-quantitative and quantitative PCRs.

Gene	Name	Sequence (5'-NNN-3') ^a	Anneal Temp (°C)	Extension time
rTetR	RR1	GAGGAACAGGAGCATCAAGTAG	58	30 s
	RR2	GCTGTAGTTCTCCAGCATCAC		
Tcf7l2	RT1	CAAGTCCTCGCTGGTCAATG	58	30 s
	RT2	CCCTCCATCTTGCCTCTTG		
Tcf7l2	RT3	GCCAAGAGGCAAGATGGAG	60	30 s
	RT4	GCGAACGAGCATCCTTGAG		
HPRT	HPRT1	GCTGGTGAAAAGGACCTCTC	60	30 s
	HPRT2	CACAGGACTAGAACACCTG		
β-actin	BACTIN1	AAGGCCAACCGTGAAAAGAT	60	30 s
	BACTIN2	CCTGTGGTACGACCAGAGGCATACA		

Anneal temperatures and extension times are given are for end-point and semi-quantitative PCR analyses.

For quantitative PCR, the same primers were used, but anneal temperature was kept constant at 60°C and extension time at 15 s. All other parameters were as recommended by the manufacturer.

Table B5: Primer pairs for cloning Tcf7l2 homology arms into pL451-TRE vector.

Arm	Sequence (5' - NNN -3') ^a	RE site	Cycles	Anneal Temp (°C)	Extension time
5'	CCACTCGAGACCCACCCTACCCGAGTTC	XhoI	30 X	70	30 s
	GGGGAATTC	EcoRI			
3'	TTTAAGCTTGTTAACCCCTTAGCTGCTG	HindIII	25 X	65	45 s
	AAAGAGCTCC	SacI			

Both homology arms were PCR amplified with KapaTaq using GC buffer, as recommended by the manufacturer for GC-rich amplicons.

Restriction enzymes sites are coloured in red.

Appendix C

Vector Maps and Sequences

ROSA26-rTetR-KRAB(Kid1) vector map

ROSA26-rTetR-KRAB(Kid1) vector sequence (1 of 5)

```

1  CTAAATTGTA AGCGTTAATA TTTGTATAA ATTGCGGTTA AATTTTGTGTT AAATCAGCTC ATTTTAAAC
71  CAATAGGCCG AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTGTTC
141 CAGTTTGGA CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA CCGTCTATCA
211 GGGCGATGGC CCACTACGTG AACCATCACC CTAATCAAGT TTTTGGGGT CGAGGTGCCG TAAAGCACTA
281 AATCGGAACC CTAAAGGAGG CCCCCGATT AGAGCTTGAC GGGGAAAGCC GGCGAACGTG GCGAGAAAGG
351 AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCAC
421 CACACCCGCC GCGCTTAATG CGCCGCTACA GGGCGCTGCC CATTGCGCAT TCAGGCTGCG CAACTGTTGG
491 GAAGGGCGAT CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG GGGATGTGCT GCAAGGCGAT
561 TAAGTTGGGT AACGCCAGGG TTTCCAGT CACGACGTTG TAAACGACG GCCAGTGAAT TGAATACGA
631 CTCACTATAG GCGCAATTGG GTACcgtaa cggcagccgg agtgcgcgc cgcgcgcgc ctcgctctgc
701 ccactgggtg gggcgaggg taggtgggtg gaggcgagct ggaagtcgag ggcggtcgg cctctggcgg
771 ggcgggggag gggaggagg gtcagcgaaa gtagctcgcg cgcgagcggc cgcaccctt cccctctctc
841 tgggggagtc gttttaccg cgcgcggcg ggcctcgtcg tctgattggc tctcggggc cagaaaactg
911 gcccttgcca ttgctcgtg ttcgtgcaag ttgagtcgat ccgcgcgcca gcgggggagg cgaggaggcg
981 ctcccaggtt ccggccctcc cctcgccccc gcgcgcgaga gtctggcgc gcgccctgc gcaacgtggc
1051 aggaagcgcg cgctgggggc ggggacgggc agtagggctg agcggctgag ggcgggtgc aagcacgttt
1121 ccgacttgag ttgctcaag aggggctgag tgagccagac ctccatcgcg cactccgggg agtgaggagg
1191 aggagcgagg gctcagttgg gctgttttgg aggcaggaag cacttgctct cccaaagtcg cctgagttg
1261 ttatcagtaa gggagctgca gtggagtagg cggggagaag gccgcaccct tctccggagg ggggagggga
1331 gtgttgcaat acctttctg gagttctctg ctgctcctg gcttctgagg accgcctggt GTACCGGGCC
1401 CCCCCTCGAA ATTctagctc tagttaatta accgtttaaa caattctgca ggaatctagt tattaatagt
1471 aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
1541 cccgcctgga tgacgcgcca acgacccccc cccattgacg tcaataatga cgtatgttcc catagtaacg
1611 ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc
1681 aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaaa tggccgcctt ggcattatgc
1751 ccagtacatg accttatggg actttcctac ttggcagtag atctacgtat tagtcacgc tattaccatg
1821 gtcgaggtga gccccacgtt ctgcttcaact ctcccatctt cccccctc cccacccca attttgtatt
1891 tatttatttt ttaattattt tgtgcagcga tggggggggg gggggggggg ggcgcgcgc caggcggggc

```


ROSA26-rTetR-KRAB(Kid1) vector sequence (2 of 5)

```

1961 ggggcggggc gaggggcggg gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc
2031 cgaaagtttc cttttatgcc gaggcggcgg cggcggcgcc cctataaaaa gcgaagcgcg cggcgggcgg
2101 ggagtcgctg cgacgctgcc ttgcgccgt gcccgctcc gccgcgcct cgcgcgcgcc gcccggtcc
2171 tgactgaccg cgttactccc acaggtgagc gggcgggacg gcccttctcc tccgggctgt aattagcgct
2241 tggtttaatg acggcttgtt tcttttctgt ggctgcgtga aagccttgag gggctccggg agggcccttt
2311 gtgcgggggg agcggctcgg ggggtgcgtg cgtgtgtgtg tgcgtgggga gcgcgcgctg cggtcccgcg
2381 ctgcccgccg gctgtgagcg ctgcggggcg gcgcggggc tttgtgcgt cgcagtgctg cgcgagggga
2451 gcgcggcccg gggcggtgcc ccgcggtgcg gggggggctg cgaggggaac aaaggctcgg tgcggggtgt
2521 gtgcgtgggg ggggtgagcag ggggtgtggg cgcgtcggtc gggctgcaac ccccctgca ccccctccc
2591 cgagttgctg agcacggccc ggcttcgggt gcggggctcc gtacggggcg tggcgcgggg ctgcgcgtgc
2661 cgggcggggg gtggcgccag gtgggggtgc cggcgggggc ggggcgcct cgggcggggg agggctcggg
2731 ggagggggcg gcgcggcccc ggagcgcccg cggtgtcga gcgcggcgca gccgcagcca ttgccttta
2801 tggtaatcgt gcgagagggc gcagggaact cctttgtccc aaatctgtgc ggagccgaaa tctgggaggg
2871 gcgcgcgcac cccctctagc gggcgcgggg cgaagcggtg cggcgccggc aggaaggaaa tggcggggga
2941 gggccttcgt gcgtgcgcgc gccgcgctcc ctttctcct ctccagctc ggggctgtcc gcggggggac
3011 ggctgccttc gggggggacg gggcagggcg gggttcggt tctggcgtgt gaccggcggc tctagagcct
3081 ctgctaacca tgttcatgcc ttcttctttt tctacagct cctgggcaac gtgctgttta ttgtctgtc
3151 tcatcathtt ggcaaagaat tgatttgata ccgcggggcc tcgacactag tgaacctctt cgagggacct
3221 aggatccacc atgtctagac tggacaagag caaagtcata aacggcgctc tgggaattact caatggagtc
3291 ggtatcgaag gcctgacgac aaggaaactc gctcaaaagc tgggagttga gcagcctacc ctgtactggc
3361 acgtgaagaa caagcgggcc ctgctcgatg ccctgccaat cgagatgctg gacaggcctc ataccactt
3431 ctgccccctg gaagcgaggt catggcaaga ctttctcgcg aacaacgcca agtcattccg ctgtgtctc
3501 ctctcacatc gcgacggggc taaagtgcgt ctcggcaccc gcccaacaga gaaacagtac gaaacctgg
3571 aaaatcagct cgcgttctg tgtcagcaag gcttctcct ggagaacgca ctgtacgctc tgtccgcgt
3641 gggccacttt acactgggct gcgtattgga ggaacaggag catcaagtag caaaagagga aagagagaca

```

ROSA26-rTetR-KRAB(Kid1) vector sequence (3 of 5)

3711 cctaccacgg atttatgcc ccactttctg agacaagcaa ttgagctggt cgaccggcag ggagccgaac
 3781 ctgccttctt ttteggcctg gaactaatca tatgtggcct ggagaaacag ctaaagtgcg aaagcggcgg
 3851 ggtcgacgtg tcgttgacat ttgaagacgt ggcctgtctc tttaccggg atgagtggaa gaagctggtt
 3921 ccctctcaga gaagcctgta ccgtgaggtg atgctggaga actacagcaa cctggcctcg ctgggattcc
 3991 cgtttaccaa accaaaagtg atctccctgc tgcagcaagg agaggacccc tggtagcgcg ccgcgactct
 4061 agatcataat cagccatacc acattttag aggttttact tgctttaaaa aacctccac acctccctc
 4131 gaacctgaaa cataaaatga atgcaattgt tgttgtaaac ttgtttattg cagcttataa tggttacaaa
 4201 taaagcaata gcatcacaaa tttcacaaat aaagcatttt tttcaactga ttctagtgtt ggtttgtcca
 4271 aactcatcaa tgtatcttaa ggatccgcg ataagggcga attctcgagt agaagatgg cgggagtctt
 4341 ctgggcaggc ttaaaggcta acctggtgtg tgggcgttgt cctgcagggg aattgaacag gtgtaaaaat
 4411 ggagggacaa gacttccac agattttcgg tttgtcggg aagttttta ataggggcaa ataaggaaaa
 4481 tgggaggata ggtagtcac tggggttta tgcagcaaaa ctacaggtta ttattgcttg tgatccgct
 4551 cggagtattt tccatcgagg tagattaaag acatgctcac ccgagtttta tactctctg cttagatcc
 4621 ttactacagt atgaaattac agtgctcgca gttagactat gtaagcagaa ttttaatcat ttttaagag
 4691 ccagtaact catatccatt tctccgctc ctctgcagc cttatcaaaa ggtattttag aacactcatt
 4761 ttagcccat tttcatttat tatactggt tatccaacc ctagacagag cattggcatt ttccctttcc
 4831 tgatcttaga agtctgatga ctcatgaac cagacagatt agttacatac accacaaac gaggctgtag
 4901 ctggggcctc aacctgcag ttcttttata actccttagt acacttttg ttgatecttt gcttgatcc
 4971 ttaattttca gtgtotatca cctctccgt caggtggtgt tccacattg ggcctattct cagtccagg
 5041 agttttacaa caatagatgt attgagaac caacctaaag cttaactttc cactcccatg aatgcctctc
 5111 tcctttttct ccattttata actgagctat taaccattaa tggtttccag gtggatgtct cctcccccaa
 5181 tattacctga tgtatcttac atattgccag gctgatattt taagacatta aaaggtatat ttcattattg
 5251 agccacatgg tattgattac tgcttactaa aattttgtca ttgtacacat ctgtaaaag tggttccttt
 5321 tggaatgcaa agttcagggt tttgttgtct ttctgacct aaggtcttgt gagcttgat tttttctatt
 5391 taagcagtgc tttctcttg actggcttga ctcatggcat tctacaggt attgctggtc taaatgtgat

ROSA26-rTetR-KRAB(Kid1) vector sequence (4 of 5)

```

5461 ttgtccaagc ttcttcagga cctataatct tgcttgactt gtagccaaac acaagtaaaa tgattaagca
5531 acaaatgtat ttgtgaagct tgggttttag gttgttgtgt tgtgtgtgct tgtgtcttat aataatacta
5601 tccaggggct ggagaggtgg ctgggagttc aagagcacag actgtctctc cagaagtcct gagttcaatt
5671 cccagcaacc acatggtggc tcacaaccat ctgtaatggg atctgatgcc ctcttctggt gtgtctgaag
5741 accacaagtg tattcacatt aaataaataa atcctccttc ttcttctttt tttttttttt aaagagaata
5811 ctgtctccag tagaatttac tgaagtaatg aaatactttg tgtttgttcc aatatggtag ccaataatca
5881 aattactctt taagcactgg aaatgttacc aaggaactaa tttttatttg aagtgttaact gtggacagag
5951 gagccataac tgcagacttg tgggatacag aagaccaatg cagactttaa tgtcttttct cttacactaa
6021 gcaataaaga aataaaaatt gaacttctag tatcctatct gtttaaactg ctagCGCGTG AAGCTTACCG

6091 CGGTGGAGCT CCAGCTTTTG TTCCCTTTAG TGAGGGTTAA TTTGAGCTT GGCCTAATCA TGGTCATAGC
6161 TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCACA CAACATACGA GCCGGAAGCA TAAAGTGTA
6231 AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCAGTCG
6301 GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC
6371 GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTGCGCTG CGGCGAGCGG TATCAGCTCA
6441 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
6511 CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG
6581 AGCATACAAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAGAT ACCAGCGCTT
6651 TCCCCTTGGG AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCGCGCTTT
6721 CTCCCTTCGG GAAGCGTGCG GCTTTCTCAT AGCTCAGCTG TAGGTATCTT CAGTTCGGTG TAGGTCGTTT
6791 GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG
6861 TCTTGTAGTC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA
6931 GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAATA CTGGTACACT AGAAGAACAG
7001 TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA
7071 ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTGTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT
7141 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGAACGA AAACCTACGT TAAGGGATTT
7211 TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTAAATTAA AAATGAAGTT TTAATCAAT
7281 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG
7351 ATCTGTCTAT TTCGTTATC CATAGTTGCC TGAATCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT
7421 TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACC GCTCCAGATT TATCAGCAAT
7491 AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCTCCAT CCAGTCTATT
7561 AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTAA ATAGTTTGCG CAACGTTGTT GCCATTGCTA
7631 CAGGCATCGT GGTGTCACGC TCGTCGTTT GTATGGCTTC ATTCAGCTCC GGTTCCTAAC GATCAAGGCG

```

ROSA26-rTetR-KRAB(Kid1) vector sequence (5 of 5)

```

7701 AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTGAGAAGT
7771 AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG
7841 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG
7911 TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT
7981 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA
8051 CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
8121 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTGAATAC TCATACTCTT CCTTTTCAA
8191 TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
8261 AACAAATAGG GGTCCGCGC ACATTTCCCC GAAAAGTGCC AC

```

pL451-TRE vector map

pL451-TRE vector sequence (1 of 2)

```

1  CTAAATTGTA AGCGTTAATA TTTTGTAAAA ATTGCGGTTA AATTTTGTGT AAATCAGCTC ATTTTAAAC
71  CAATAGGCCG AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTGTTC
141 CAGTTTGGA CAAGAGTCCA CTATTAAGA ACGTGGACTC CAACGTCAA GGGCGAAAAA CCGTCTATCA
211 GGGCGATGGC CCACTACGTG AACCATCACC CTAATCAAGT TTTTGGGGT CGAGGTGCCG TAAAGCACTA
281 AATCGGAACC CTAAAGGGAG CCCCCGATT AGAGCTTGAC GGGGAAAGCC GGCGAACGTG GCGAGAAAGG
351 AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCCAC
421 CACACCCGCC GCGCTTAATG CGCCGCTACA GGGCGCGTCC CATTGCGCAT TCAGGCTGCG CAACTGTTGG
491 GAAGGGCGAT CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG GGGATGTGCT GCAAGGCGAT
561 TAAGTTGGGT AACGCCAGGG TTTTCCAGT CACGACGTTG TAAACGACG GCCAGTGAAT TGTAAACGA
                                     KpnI      XhoI      NotI      EcoRI
631 CTCCTATAG GGCGAATTGG GTACCGGGCC CCCCTCGAG agctgcgccc gctgcaGAAT TCCATCGATT
                                     MluI      HindIII  SacII      SacI
981 TGTGTCCTT TATAAGACGC GTGAAGCTTA CCGCGGTG GAGCTCCAGC TTTGTTCC TTAGTGAGG
1049 GTTAATTTTC AGCTTGGCGT AATCATGCTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT
1119 CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAGCCCT GGGGTGCTTA ATGAGTGAGC TAACACACAT
1189 TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG
1259 CCAACGCGCG GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC
1329 TCGGTGCTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG
1399 GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT
1469 GCTGGCGTTT TTCCATAGGC TCCGCCCCC TGACGAGCAT CAAAAAATC GAGCTCAAG TCAGAGGTGG
1539 CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCC CTGGAAGCTC CCTCGTGC GC TCTCCTGTT
1609 CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC
1679 ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCGTT
1749 CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC
1819 CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA
1889 GTGTGGCCTT AACTACGGCT AACTAGAAG AACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC
1959 TTCGGAAAAA GAGTTGGTAG CTCTGTATCC GGCAAAACAA CCACCGCTGG TAGCGGTGGT TTTTGTGTTT
2029 GCAAGCAGCA GATTACGCGC AGAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTCTA CGGGGTCTGA
2099 CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTACCTAG
2169 ATCCTTTTAA ATTAAAAATG AAGTTTAA TCAATCTAAA GTATATATGA GTAACTTGG TCTGACAGTT
2239 ACCAATGCTT AATCAGTGA GCACCTATCT CAGCGATCTG TCTATTTCTG TCATCCATAG TTGCCTGACT
2309 CCCGTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCA GTGCTGCAAT GATACCGCA
2379 GACCCACGCT CACCGGCTCC AGATTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG

```

pL451-TRE vector sequence (2 of 2)

```
2449  GTCTGCAAC TTTATCCGCC TCCATCCAGT CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC
2519  AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC GTTTGGTATG
2589  GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG
2659  TTAGCTCCTT CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC
2729  AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCCTTTCTG TGA CTGGTGA G TACTCAACC
2799  AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTGCCCCGGC GTC AATACGG GATAATACCG
2869  CGCCACATAG CAGAACTTTA AAAGTGCTCA TCATTGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT
2939  CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT GCACCCAAC TATCTTCAGC ATCTTTTACT
3009  TTCACCAGCG TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC
3079  GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT
3149  GAGCGGATAC ATATTTGAAT GTATTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA
3219  GTGCCAC
```

pMYS-Nup98HoxB4-IRES-Puro vector map

pMYS-Nup98HoxB4-IRES-Puro vector sequence (1 of 4)

```

1  ctgcataatg aaagacccca cctgtagggt tggcaagcta gcttaagtaa cgccattttg caaggcatgg
71  aaaaatacat aactgagaat agaaaagtgc agatcaaggt caggaacaga tggaaacagct gaatatgggc
141  caaagcggat atctgtggta agcagttcct gccccggctc agggccaaga acagatggaa cagctgaata
211  tgggcaaac aggataatctg tggtaagcag ttctcgcccc ggctcagggc caagaacaga tgggtccccag
281  atgcggtcca gccctcagca gtttctagag aaccatcaga tgtttccagg gtgccccaaag gacctgaat
351  gaccctgtgc cttatttgaa ctaaccaatc agttcgcttc tcgcttctgt tcgcgcgctt ctgctccccg
421  agctcaataa aagagcccac aaccctcac tcggggcgcc agtcctccga ttgactgagt cgccccggta
491  cccgtattcc caataaagcc tcttgctggt tgcacccgaa tcgtggactc gctgacctt gggagggtct
561  cctcagattg attgactgcc cacctcgggg gtctttcatt tggaggttcc accgagattt ggagacccca
631  gccacgggac caccgacccc ccgcgcggga ggtaagctgg ccagcggctg ttctgtgtct gtctctgtct
701  ttgtgcgtgt ttgtgcgggc atctaagtgt tgcgcctgct tctgtactag ttagctaact agctctgtat
771  ctggcggacc cgtggtggaa ctgacgaggt cggaacaccc ggccgcaacc ctgggagacg tcccagggac
841  ttccggggcc gttttgtggg cccgacctga gtccaaaaat cccgatcggt ttggactctt tgggtcaccc
911  cccttagagg agggatatgt ggttctggta ggagacgaga acctaaaaca gttccccctt ccgtctgaat
981  ttttgccttc ggtttgggac cgaagccgct ccgcgcgtct tgtctgctgc agcatcgctt tgtgtgtct
1051  ctgtctgact gtgtttctgt atttgtctga aaatatgggc ccgggccaga ctgttaccac tcccttaagt
1121  ttgaccttag gtcactggaa agatgtcgag cggatcgctc acaaccagtc ggtagatgac aagaagagac
1191  gttgggttac cttctgctct gcagaatggc caacctttaa cgtcggatgg ccgcgagacg gcaccttaa
1261  ccgagacctc atcaccacag ttaagatcaa ggtcttttca cctggcccgc atggacaccc agaccaggtc
1331  ccctacatcg tgacctggga agccttggtt ttgaccccc ctccctgggt caagcccttt gtacacccta
1401  agcctccgcc tctcttctct ccacccgccc cgtctctccc ccttgaacct cctcgttcga ccccgctctg
1471  atcctccctt tatccagccc tcaactcttc tctaggcgcc cccatatggc catatgagat cttatatggg
1541  gcacccccgc cccttgtaaa cttccctgac cctgacatga caagagttac taacagcccc tctctccaag
1611  ctcaattaca ggctctctac ttagtcagc acgaagtctg gagacctctg gcggcagcct accaagaaca
1681  actggaccga ccggtggtac ctcaccctta ccgagtcggc gacacaggtg gggtcgcgcg acaccagact
1751  aagaacctag aacctcgctg gaaaggacct tacacagtcg tgctgaccac ccccaccgcc ctcaaagtag
1821  acggcatcgc agcttgata cacgcgccc acgtgaaggc tgccgacccc gggggtggac catctctag
1891  actgcgggat ccaagatgtt taacaaatca ttggaacac cctttggggg tggcacaggt ggctttggca
1961  caacttcaac atttgacag aatactggct ttggcactac tagtgagggg gcatttgaa catctgcatt
2031  tggttctagc aacaatactg gaggcctctt tggaaattca cagactaaac caggaggatt gtttgaacc

```

pMYS-Nup98HoxB4-IRES-Puro vector sequence (2 of 4)

```

2101 agttcattta gccagccagc tacctccaca agcactggct ttgggttttg tacgtcaaca ggaacagcaa
2171 ataccttggt tggaactgca agcacaggga ccagtctctt ctcaccccaa acaatgcct ttgcacaaaa
2241 taaaccaact ggctttggca attttgaac cagtactagc agtggaggac tctttggaac cacaaatacc
2311 acctctaata cttttggcag cacatctggc tccctctttg ggccaagtag ttttacagct gctcctactg
2381 ggactactat taaatttaac cctccaactg gtacagatac tatggcmeta gctggagtta gcaactaat
2451 aagtaccaag caccagtgtt ttactgctat gaaagaatat gaaagcaagt cactagagga acttcgttta
2521 gaggattatc aggctaacag gaagggccca cagaaccagg tgggagcagg taccacaact ggcttgtttg
2591 ggtcttctcc agccacttcc agcgcaacag gactcttcag ctctccacc actaattcag gctttgcata
2661 ttggtcagaac aaaactgcct ttggaactag tacaactgga tttggaaca atccaggtag tctcttggc
2731 caacagaatc agcagactac cagcctcttc agcaaacat ttggccaggc tacaaccacc cagaacactg
2801 gcttttctct ttgtaatacc agcaccatag gacagccaag caccaacacc atgggattat ttggagtaac
2871 ccaagcctca cagcctggag gtcttttttg gacagctaca aacaccagca ctgggacagc atttgaaca
2941 ggaacaggtc tctttgggca gaccaatact ggatttgggt ctgttgggtc gacctgttt ggcaataaca
3011 agcttactac atttgaagc agcacaacca gtgcaccttc atttggtag accagtggcg ggtcttttg
3081 ttttggcaca aataccagtg ggaatagtat ttttgaagt aaaccagcac ctgggactct ttggaactgg
3151 cttgtgtcag gatttgaac agctcttggg gctggacagg catctttgtt ttggaacaac caacctaa
3221 ttggagggcc tcttggtaga ggagcctttg gggccctgg atttaatact acgacagcca ctttgggctt
3291 ttggagcccc caggccccag tagaattcgt aaacccaat tacgcccggc gggagcccaa gcgctctcg
3361 accgcctaca cgcgccagca ggtcttggag ctggagaagg aatttacta caaccgtac ctgacacggc
3431 gccggagggt ggagatcgcc cagcgctctt gcctctccga gcgccagatc aagatctggt tccagaaccg
3501 gcgcatgaag tggaaaaaag accacaagtt gcccaacacc aagatccgct cgggtgtgtc ggcaggctca
3571 gccggagggt cccctggcgg gcccaatgga ggcgcccgcg cgctctagct cgaggggcgg cgcgcccgcg
3641 ccgctacgta aattccgccc ctctccctcc ccccccccta acgttactgg ccgaagccgc ttggaataag
3711 gccggtgtgc gtttgtctat atgttatttt ccaccatatt gccgtctttt ggcaatgtga gggcccgaa
3781 acctggccct gtcttcttga cgagcattcc taggggtctt tccctctctg ccaaggaat gcaaggctct
3851 ttgaatgtcg tgaagggaag agttctctct gaagcttctt gaagacaaac aacgtctgta gcgaccttt
3921 gcaggcagcg gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc
3991 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtaa atggctctcc
4061 tcaagcgtat tcaacaagg gctgaaggat gcccagaagg taccctattg tatgggatct gatctggggc
4131 ctcggtgcac atgctttaca tgtgtttagt cgagggttaa aaacgtctag gccccccgaa ccacggggac

```

pMYS-Nup98HoxB4-IRES-Puro vector sequence (3 of 4)

```

4201 gtggttttcc ttgaaaaac acgatgataa gcttgccaca accatggccg agtacaagcc cacggtgcgc
4271 ctcgccaccc ggcagcagct ccccagggcc gtacgcaccc tcgcccgcgc gttcgccgac taccgcccca
4341 cgcgccacac cgtcgatccg gaccgccaca tcgagcgggt caccgagctg caagaactct tcctcacgcg
4411 cgtcgggctc gacatcggca aggtgtgggt cgcggacgac ggccgcccgc tggcggtctg gaccacgcgc
4481 gagagcgtcg aagcgggggc ggtgttcgcc gagatcggcc cgcgcattgc cgagttgagc ggttcccggc
4551 tggccgcgca gcaacagatg gaaggcctcc tggcgccgca ccggcccaag gagcccgctg ggttcctggc
4621 caccgtcggc gtctcgcccc accaccaggg caagggtctg ggcagcgcgc tcgtgctccc cggagtggag
4691 gcggccgagc gcgccggggt gccgccttc ctggagacct ccgcgccccg caacctcccc ttctacgagc
4761 ggctcggctt caccgtcacc gccgacgtcg aggtgccccg aggcacgcgc acctggtgca tgaccgcgaa
4831 gcccggtgcc tgactcgacc ggcccgccag cacagtgggg gggggcccgc gattagtcca atttgttaa
4901 gacaggatat cagtgtgtcca ggctctagtt ttgactcaac aatatcacca gctgaagcct atagagtacg
4971 agccatagat agaataaaag attttattta gtctccagaa aaagggggga atgaagacc ccacctgtag
5041 gtttggaagc ctagtctaac taagccattt tgcaaggcat ggaaaaatac ataactgaga atagagaagt
5111 tcagatcaag gtttagaaca gagagacagg agaatatggg ccaaacagga tatctgtggt aagcagttcc
5181 tgccccggtc cagggccaaag aacagttgga acagcagaat atggggccaa caggatatct gtggaagca
5251 gttcctgccc cggctcaggg ccaagaacag atggtcccca gatgcggtcc cgccctcagc agtttctaga
5321 gaaccatcag atgtttccag ggtgccccaa ggacctgaaa tgacctgtg ccttatttga actaaccaat
5391 cagttcgtct ctgcgttctg ttcgcgcgct totgctcccc gagtcaata aaagagccca caaccctca
5461 ctcgccgcgc cagtcctccg atagactcgc tcgcccgggt acccgtgtat ccaataaacc ctcttgagc
5531 tgcatccgac ttgtggtctc gctgttcctt gggagggtct cctctgagtg attgactacc cgtcagcggg
5601 ggtctttcat tctgcattaa tgaatcgcc aacgcgcggg gagaggcggt ttgcgtattg ggcgctctc
5671 cgcttcctcg ctcaactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcaactcaaag
5741 gcggttaatac ggttatccac agaatacagg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa
5811 aggccaggaa ccgtaaaaaa gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca
5881 caaaaatcga cgtcgaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc gtttcccctt
5951 ggaagctccc tcgtgcgctc tctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt
6021 cgggaagcgt gccgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa
6091 gctgggctgt gtgcacgaac cccccgttca gccgcaccgc tgcgccttat ccggttaacta tcgtcttgag
6161 tccaaccgga taagacacga ctatcgcca ctggcagcag ccaactggtaa caggattagc agagcagagt
6231 atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg
6301 tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct ctgtatccgg caaacaacc
6371 accgctggta gcggtggttt tttgttttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag
6441 atcctttgat cttttctacg ggtctgacg ctcagtggaa cgaaaactca cgtaaggga ttttggtcat
6511 gagattatca aaaagatct tcacctagat ccttttgcg ccggccgcaa atcaatctaa agtatatatg
6581 agtaaaactg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttgc
6651 ttcattcata gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc
6721 agtgctgcaa tgataccgag agaccacgc tcaccggctc cagatttatc agcaataaac cagccagccg

```

pMYS-Nup98HoxB4-IRES-Puro vector sequence (4 of 4)

```

6791 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga
6861 agctagagta agtagttcgc cagttaatag ttgcgcaac gttgttgcca ttgctacagg catcgtggtg
6931 tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgac aaggcgagtt acatgatccc
7001 ccatgttggtg caaaaaagcg gttagctcct tcggtcctcc gatcgttggtc agaagtaagt tggccgcagt
7071 gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct
7141 gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccg
7211 cgtcaacacg ggataatacc gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgtttctc
7281 ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcg tgcaccaac
7351 tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa
7421 aaaagggaat aaggcgaca cggaatggt gaatactcat actcttcctt tttcaatatt attgaagcat
7491 ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt
7561 ccgcgcacat ttc

```

pMYS-TRE-IRES-GFP vector map

pMYS-TRE-IRES-GFP vector sequence (1 of 4)

```

1  ctgcataatg aaagacccca cctgtaggtt tggcaagcta gcttaagtaa cgccattttg caaggcatgg
71  aaaaatacat aactgagaat agaaaagttc agatcaaggt caggaacaga tggaaacagct gaatatgggc
141  caaagcggat atctgtggta agcagttcct gccccggctc agggccaaga acagatggaa cagctgaata
211  tgggccaaac aggatatctg tggtaagcag ttctgcccc ggctcagggc caagaacaga tgggtccccag
281  atgcggtcca gccctcagca gtttctagag aaccatcaga tgtttccagg gtgccccaaag gacctgaaat
351  gaccctgtgc cttatttgaa ctaaccaate agttcgcttc tcgcttctgt tcgcgcgctt ctgctccccg
421  agctcaataa aagagcccac aaccctcac tcggggcgcc agtcctccga ttgactgagt cggccgggta
491  cccgtatttc caataaagcc tcttgctggt tgcacccgaa tcgtggactc gctgacccct gggaggggtc
561  cctcagattg attgactgcc caectcgggg gtctttcatt tggaggttcc accgagattt ggagacccca
631  gccagggac caccgacccc cccgcccggg ggttaagctg ccagcggctg ttctgtgtct gtctctgtct
701  ttgtgcgtgt ttgtgccgcg atctaattgt tcgcctcgcg tctgtactag ttagctaaat agctctgtat
771  ctggcggacc cgtggtggaa ctgacgagtt cggaacaccc ggccgcaacc ctgggagacg tcccagggac
841  ttccggggcc gtttttgtgg cccgacctga gtccaaaaat cccgatcggt ttggaactct tgggtgcacc
911  cccttagagg agggatatgt gtttctggta ggagacgaga acctaaaaca gttcccgcct cgtctgaat
981  ttttgccttc ggtttgggac cgaagccgcg ccgcgcgtct tgtctgtgc agcatcgctc tgtgtgtct
1051  ctgtctgact gtgtttctgt atttgtctga aaatatgggc cggggccaga ctgttaccac tcccttaagt
1121  ttgaccttag gtcactggaa agatgtcgag cggatcgctc acaaccagtc ggtagatgac aagaagagac
1191  gttgggttac cttctgtctt gcagaatggc caacctttaa cgtcggatgg ccgcgagacg gcacctttaa
1261  ccgagacctc atcaccagc ttaagatcaa ggtcttttca cctggcccgc atggacaccc agaccaggtc
1331  ccctacatcg tgacctggga agccttggtt ttgaccccc ctccctgggt caagcccttt gtacacccta
1401  agcctccgct tcctcttctt ccatccgccc cgtctctccc cettgaacct cctcgttoga cccgcctcgt
1471  atcctccctt tatccagccc tcaactcttc tctagcgccc cccatatggc catatgagat cttatatggg
1541  gcacccccgc cccttgtaaa ctccctgac cctgacatga caagagttac taacagcccc tctctccaag
1611  ctcaactaca ggctctctac ttagtccagc acgaagtctg gagacctctg gcggcagcct accaagaaca
1681  actggaccga ccggtggtac ctaccctta ccgagtcggc gacacagtgt gggtcgcgcg acaccagact

```

pMYS-TRE-IRES-GFP vector sequence (2 of 4)

```

1751 aagaacctag aacctcgctg gaaaggacct tacacagtcc tgctgaccac cccaccgcc ctcaaagtag
1821 acggcatcgc agcttgata cagcgcgccc acgtgaaggc tgccgacccc gggggtggac catcctctag
1891 actgccggat cccagtgtgg tggtagcgga attccatcga ttgcgaatt gtcgagttta ctccctatac
1961 gtgatagaga acgtatgtcg agtttactcc ctatcagtga tagagaacga tgcgagttt actccctatc
2031 agtgatagag aacgtatgtc gagtttactc cctatcagtg atagagaacg tatgtcgagt ttactcccta
2101 tcagtgatag agaacgtatg tcgagtttat ccctatcagt gatagagaac gtatgtcgag ttactccctt
2171 atcagtgata gagaacgtat gtcgaggtag gcatagatcc aatgttgtcc ttataagac gcgtaagct
2241 taccgcggcc gctacgtaaa ttccgcccct ctccctcccc ccccctaacc gttactggcc gaagccgctt
2311 ggaataaggc cgggtgtcgt ttgtctatat gttattttcc accatattgc cgtcttttgg caatgtgagg
2381 gcccggaaac ctggccctgt cttcttgacg agcattccta ggggtcttcc ccctctcgcc aaaggaatgc
2451 aaggctctgt gaatgtcgtg aaggaagcag ttcctctgga agcttcttga agacaaacaa cgtctgtagc
2521 gaccctttgc aggcagcgga acccccacc tggcgacagg tgcctctgcg gccaaaagcc acgtgtataa
2591 gatacacctg caaaggcggc acaacccacg tgccacgttg tgagttggat agttgtggaa agagtcaaat
2661 ggctctcttc aagcgtatto acaaggggc tgaaggatgc ccagaaggtg cccattgta tgggatctga
2731 tctggggcct cgggtgcacat gctttacatg tgtttagtcg aggttaaaaa acgtctaggg cccccaacc
2801 acggggacgt ggttttctt tgaaaaaac gatgataagc ttgccacaac catggtgagc aagggcgagg
2871 agctgttcac cggggtgtg cccatcctgg tcgagctgga cggcgacgta aacggccaca agttcagcgt
2941 gtccggcgag ggcgagggcg atgccaccta cggcaagctg accctgaagt tcatctgcac caccggcaag
3011 ctgcccgctg cctggccacc cctcgtgacc accctgacct acggcgtgca gtgcttcagc cgtacccccg
3081 accacatgaa gcagcagcag ttcttcaagt ccgccatgcc cgaaggctac gtccaggagc gcaccatctt
3151 cttcaaggac gacggcaact acaagaccgg cgccgaggtg aagttcgagg gcgacaccct ggtgaaccgc
3221 atcgagctga agggcatcga cttcaaggag gacggcaaca tctgggggca caagctggag tacaactaca

```

1► M V S K G E
 7► E L F T G V V P I L V E L D G D V N G H K F S V
 30► S G E G E G D A T Y G K L T L K F I C T T G K
 54► L P V P W P T L V T T L T Y G V Q C F S R Y P
 77► D H M K Q H D F F K S A M P E G Y V Q E R T I F
 100► F K D D G N Y K T R A E V K F E G D T L V N R
 124► I E L K G I D F K E D G N I L G H K L E Y N Y

pMYS-TRE-IRES-GFP vector sequence (3 of 4)

```

3291 acagccacaa cgtctatatc atggccgaca agcagaagaa cggcatcaag gtgaacttca agatccgcga
147▶ N S H N V Y I M A D K Q K N G I K V N F K I R H
3361 caacatcgag gacggcagcg tgcagctcgc cgaccactac cagcagaaca ccccatcgcg cgacggcccc
170▶ N I E D G S V Q L A D H Y Q Q N T P I G D G P
3431 gtgctgctgc ccgacaacca ctacctgagc acccagtcgc cctgagcaa agaccccaac gagaagcgcg
194▶ V L L P D N H Y L S T Q S A L S K D P N E K R
3501 atcacatggt cctgctggag ttctgtacgc ccgccgggat cactctcggc atggacgagc tgtacaagta
217▶ D H M V L L E F V T A A G I T L G M D E L Y K .
3571 aagcggccca gccacagtgg tcgaccgggc cgccagcaca gtgggggggg gcccgcgatt agtccaattt
240▶
3641 gttaaagaca ggatatcagt ggtccaggct ctagtcttga ctcaacaata tcaccagctg aagcctatag
3711 agtacgagcc atagatagaa taaaagattt tatttagtct ccagaaaaag gggggaatga aagacccac
3781 ctgtaggttt ggcaagctag cttaagtaag ccattttgca aggcattgaa aaatacataa ctgagaatag
3851 agaagttcag atcaaggtta ggaacagaga gacaggagaa tatgggccaa acaggatatc tgtggttaagc
3921 agttcctgcc ccggctcagg gccagaaca gttggaacag cagaatatgg gccaaacagg atatctgtgg
3991 taagcagttc ctgccccggc tcagggccaa gaacagatgg tccccagatg cggtcgccgc ctccagcagtt
4061 tctagagaac catcagatgt ttccagggtg cccaaggac ctgaaatgac cctgtgcctt atttgaacta
4131 accaatcagt tcgtctctcg ctctgttctg cgcgcttctg ctccccgagc tcaataaaag agccccacaac
4201 ccctcactcg gcgcgccagt cctccgatag actgcgtcgc ccgggtaccc gtgtatccaa taaacacctt
4271 tgcagttgca tccgacttgt ggtctcgtcg ttccctggga ggtctctctc tgagtgattg actaccgcgc
4341 agcgggggtc ttctattctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg
4411 ctcttccgct tcctcgtcga ctgactcgtc gcgctcggtc gttcggctgc ggcgagcggt atcagctcac
4481 tcaaagcgcg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc
4551 agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggtcccgcc ccctgacga
4621 gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt
4691 cccctggaa gtcctctcgt gcgctctcct gttccgaccc tgcgcttac cggataacct tccgcctttc
4761 tcccttcggg aagcgtggcg ctttctcata gtcacgctg taggtatctc agttcggtgt aggtcgttcg
4831 ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ctttatccgg taactatcgt
4901 cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacag attagcagag
4971 cgagggtatg aggcggtgct acagagtctc tgaagtggg gcctaactac ggctacacta gaaggacagt
5041 atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa
5111 caaaccaccg ctggtagcgg tggttttttt gtttgaagc agcagattac gcgcagaaaa aaaggatctc
5181 aagaagatcc tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatctt
5251 ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttgcggccgg ccgcaaatca atctaaagta

```


pMYS-TRE-IRES-GFP vector sequence (4 of 4)

```

5321 tatatgagta aacttgggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct

5391 atttcgttca tccatagttg cctgaactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct

5461 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc

5531 cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg

5601 ccgggaagct agagtaagta gttcgccagt taatagtgtg cgcaacgttg ttgccattgc tacaggcatc

5671 gtggtgtcac gctcgtcgtt tggatatggt tcattcagct ccggttccca acgatcaagg cgagttacat

5741 gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tctccgatac gttgtcagaa gtaagttggc

5811 cgcagtggtt tcaactatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc

5881 ttttctgtga ctgggtagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt

5951 gccccgcgtc aacacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg

6021 ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca

6091 cccaactgat cttcagcacc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg

6161 ccgcaaaaaa gggaataaagg gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg

6231 aagcatttat cagggttatt gtotcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
6301 ggggttccgc gcacatttc

```

Appendix D

Permission Agreements

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Aug 17, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4411410944816
License date	Aug 17, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Molecular Cell Biology
Licensed Content Title	Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration
Licensed Content Author	Nick Barker
Licensed Content Date	Dec 11, 2013
Licensed Content Volume	15
Licensed Content Issue	1
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	An inducible and reversible genetics platform for novel drug target discovery
Instructor name	Prof. Klaus Karjalainen
Institution name	Nanyang Technological Institute
Expected presentation date	Aug 2018
Portions	Figure 1
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Billing Type	Invoice
Billing Address	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Attn: Mei Ling Leong
Total	0.00 USD

<https://s100.copyright.com/AppDispatchServlet>

1/3

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Aug 17, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4411440580256
License date	Aug 17, 2018
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	Probing the impact of chromatin conformation on genome editing tools
Licensed content author	Chen, Xiaoyu; Rinsma, Marrit
Licensed content date	Jun 8, 2016
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	An inducible and reversible genetics platform for novel drug target discovery
Publisher of your work	Nanyang Technological Institute
Expected publication date	Aug 2018
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	An inducible and reversible genetics platform for novel drug target discovery
Instructor name	Prof. Klaus Karjalainen
Institution name	Nanyang Technological Institute
Expected presentation date	Aug 2018
Portions	Figure 1A
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Publisher Tax ID	GB125506730
Billing Type	Invoice
Billing Address	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Attn: Mei Ling Leong
Total	0.00 USD
Terms and Conditions	

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Aug 17, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4411440439603
License date	Aug 17, 2018
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Transcriptional activation by tetracyclines in mammalian cells
Licensed Content Author	M Gossen,S Freundlieb,G Bender,G Muller,W Hillen,H Bujard
Licensed Content Date	Jun 23, 1995
Licensed Content Volume	268
Licensed Content Issue	5218
Volume number	268
Issue number	5218
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	
Title of your thesis / dissertation	An inducible and reversible genetics platform for novel drug target discovery
Expected completion date	Aug 2018
Estimated size(pages)	1
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Billing Type	Invoice
Billing Address	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Attn: Mei Ling Leong
Total	0.00 USD

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Aug 17, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4411431066048
License date	Aug 17, 2018
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting
Licensed Content Author	H Gu,JD Marth,PC Orban,H Mossmann,K Rajewsky
Licensed Content Date	Jul 1, 1994
Licensed Content Volume	265
Licensed Content Issue	5168
Volume number	265
Issue number	5168
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	
Title of your thesis / dissertation	An inducible and reversible genetics platform for novel drug target discovery
Expected completion date	Aug 2018
Estimated size(pages)	1
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Billing Type	Invoice
Billing Address	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Attn: Mei Ling Leong
Total	0.00 USD

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS

ELSEVIER LICENSE TERMS AND CONDITIONS

Aug 19, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4412480574013
License date	Aug 19, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Biotechnology
Licensed Content Title	Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit
Licensed Content Author	Ioannis Mougiakos, Elleke F. Bosma, Willem M. de Vos, Richard van Kranenburg, John van der Oost
Licensed Content Date	Jul 1, 2016
Licensed Content Volume	34
Licensed Content Issue	7
Licensed Content Pages	13
Start Page	575
End Page	587
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	An inducible and reversible genetics platform for novel drug target discovery
Publisher of new work	Nanyang Technological Institute
Author of new work	Prof. Klaus Karjalainen
Expected completion date	Aug 2018
Estimated size (number of pages)	1
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Aug 19, 2018

This Agreement between Nanyang Technological Institute -- Mei Ling Leong ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4412410325137
License date	Aug 19, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Protocols
Licensed Content Title	Genome engineering using the CRISPR-Cas9 system
Licensed Content Author	F Ann Ran, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott et al.
Licensed Content Date	Oct 24, 2013
Licensed Content Volume	8
Licensed Content Issue	11
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	An inducible and reversible genetics platform for novel drug target discovery
Instructor name	Prof. Klaus Karjalainen
Institution name	Nanyang Technological Institute
Expected presentation date	Aug 2018
Portions	Figure 1
Requestor Location	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Singapore Attn: Mei Ling Leong
Billing Type	Invoice
Billing Address	Nanyang Technological Institute 60 Nanyang Drive Singapore, Singapore 637551 Attn: Mei Ling Leong
Total	0.00 USD

<https://s100.copyright.com/AppDispatchServlet>

1/3